DOCUMENT RESUME

ED 179 959 CS 205 293

AUTHOR	Evans, Peter J. A.
TITLE	Evaluation of Writing in Ontario: Grades 8, 12, and 13: Analysis and Comparison of Writing Performance of Grade 8 Students in the 1977 Ontario Institute for Studies in Education Intermediate Evaluation Project and of Grades 12 and 13 Students in the Interface Study, 1976. Review and Evaluation Bulletins, Volume 1, Number 2, 1979.
INSTITUTION	Ontario Inst. for Studies in Education, Toronto.
SPONS AGENCY	Ontario Dept. of Education, Toronto.
PUB DATE	79
NOTE	118p.
AVAILABLE FROM	The Ontario Institute for Studies in Education, Publications Sales, 252 Bloor St. West, Toronto, Ontario, Canada M5S 1V6
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Composition (Literary); *Educational Research; *English Instruction; Error Analysis (Language); Evaluation; Foreign Countries; *Language Development; Secondary Education; *Student Evaluation; *Writing Skills

ABSTRACT

This condensation of an original report of the same title describes an analysis and comparison of writing performances of students in two studies: grade 8 students in the 1977 Ontario Institute for Studies in Education (OISE) Intermediate Evaluation Project and grade 12 and 13 students in the Interface Study, 1976. The chapters give the purposes of the study, discuss the evaluation of writing in the Interface Study and in the OISE Intermediate Evaluation Project, describe writing performances at each of the grade levels studied, describe the development in writing ability of students from grades 8 through 12, and describe a program for writing evaluation within a school system. The report concludes with a chapter that describes a writing evaluation program for a school or system at the intermediate or senior level. An appendix includes some examples of student writing with comments from an evaluator. (TJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

REVIEW AND EVALUATION BULLETINS

Volume 1, Number 2
1979

Evaluation of Writing in Ontario: Grades 8, 12, and 13

ED179959

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
The Minister of Education, Ontario
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Ministry of Education
Ontario

CS205293

EVALUATION OF WRITING IN ONTARIO: GRADES 8, 12, AND 13

Analysis and Comparison of Writing Performance of Grade 8 Students in the 1977 Ontario Institute for Studies in Education Intermediate Evaluation Project and of Grades 12 and 13 Students in the Interface Study, 1976

PETER J. A. EVANS, Principal Investigator

This research project was funded under contract
by the Ministry of Education, Ontario

© The Minister of Education, Ontario, 1979
Queen's Park
Toronto, Ontario

This study reflects the views of the author(s) and not necessarily those of the Ministry of Education.

Printed by The Ontario Institute for Studies in Education

MINISTRY OF EDUCATION, ONTARIO CATALOGUING IN PUBLICATION DATA

Evans, Peter J. A.
The evaluation of writing in Ontario : grades 8, 12, and 13 / Peter J. A. Evans assisted by Marion Marsh
(Review and evaluation bulletins / Ontario Ministry of Education ; v.1, no.2)
"Analysis and comparison of writing performance of grade 8 students in the 1977 Ontario Institute for Studies in Education Intermediate Evaluation Project and of Grade 12-13 students in the Interface study, 1976."
"This monograph is based on a report with the same title presented to the Ministry of Education in April 1973."
Study funded by the Ministry of Education, Ontario.

1. English composition test. 2. Creative writing (Elementary education). 3. Creative writing (Secondary education).
I. Marsh, Marion. II. Ontario. Ministry of Education. III. Title.
IV. Title: Intermediate evaluation project. V.: Interface study.
VI. Series: Ontario. Ministry of Education. Review and evaluation bulletin ; v.1, no.2.
ON01454 808.072

Additional copies may be ordered from:

The Ontario Institute for Studies in Education
Publications Sales
252 Bloor Street West
Toronto, Ontario
M5S 1V6

or from:

The Ontario Government Bookstore
880 Bay Street
Toronto, Ontario
M7A 1L2

Record Number ON01454
1 2 3 4 5 XX 78 28 18 08 97

Contents

1 Purposes of the Study/*1*

2 The Evaluation of Writing in the Interface Study and in the OISE Intermediate Evaluation Project/*6*

3 Performance: Grade 8 Writing/*19*

4 Performance: Grades 12 and 13/*48*

5 Development in Writing Ability: Grades 8-12/*63*

6 A Program for Writing Evaluation within a School System/*67*

Bibliography/*84*

Appendix/*87*

Preface

This monograph is in some respects a condensation of the original report, with the same title, prepared for the Research and Evaluation Branch of the Ontario Ministry of Education in April 1978, which was funded under contract by the Ministry of Education. The original report is available from the Ministry of Education, Ontario, and from the regional offices of the Ministry.

I have kept as much as possible of the basic data on student performance which forms the basis of that report, eliminating a good deal of the information concerning the framework of the Interface Study as that information is available in the published reports of that study.

The original report *The Evaluation of Writing in Ontario: Grades 8, 12, and 13* contains, as chapter 6, an extension of the comparison of writing performance to research studies performed elsewhere, and in particular to the reports of the American National Assessment of Educational Progress concerning the writing performance of 13- and 17-year-olds. That entire chapter was prepared and written by Marion Marsh. Consultants and researchers will find it useful. I encourage those wishing to pursue the analysis of writing performance further to consult the full-length report and in particular that chapter.

Chapters 1 to 5 of this monograph are edited versions of the longer report. The original report concludes with two chapters exploring in a tentative way the program implications of our findings and alter-

native approaches to evaluation. Some of those ideas are carried across into the concluding chapter of this monograph, but rewritten in a form I hope will be more useful to teachers and consultants in planning writing evaluation strategies for a school or system at the Intermediate or Senior level.

That is the main purpose in this monograph presentation: once you, as teachers, consultants, or administrators, have examined the data revealed by two reasonably successful studies of student writing (chapters 2 to 5), we hope you will be encouraged to undertake a systematic approach to writing evaluation in your system. Consequently, the concluding chapter is somewhat more dogmatic in tone than those concerned with the subtler questions of writing evaluation might expect or wish. I know from two years' experience now with the Intermediate Writing Evaluation Project that teachers generally welcome a coherent approach, program related, to writing evaluation. What they need is direct help and advice as to how it can be managed effectively. So I have attempted to describe procedures that *do work*.

In the original report there are a number of appendices, which provide our specific instructions to scorers for error counts and syntactic maturity measures and full tables of counts and scores on each essay. I have included here only the appendix, which provides samples with comments of the grades 12-13 essays from the Interface Report.

Peter Evans

Acknowledgments

In the course of the 1976 Interface Study funded jointly by the Ontario Ministry of Education and Ministry of Colleges and Universities, we collected and scored almost 1600 essays of grades 12 and 13 students, a careful sampling of schools and students in the province. In the spring of 1977, over 6000 students in grades 7 and 8 wrote a pair of assignments in context of the Ontario Institute for Studies in Education's Intermediate Evaluation Project. About eighteen Ontario school systems were involved. A sample of these compositions was subsequently analysed (error counts and analytic scoring) as part of that project and the report (informal) made available to participating schools.

For the first time in Ontario, we had a considerable amount of data on the characteristics of student writing at these two levels. As well, in both studies we learned a great deal about the handling, scoring, and analysis of student writing. It seemed a pity to leave matters there when so much of the information could be helpful to teachers in their own analysis and evaluation of student writing and evaluation of writing programs.

The Ministry of Education, therefore, provided funds to extend the analysis of both the Intermediate and Interface writings and to make such comparisons between these two levels as might prove useful.

Our acknowledgements must include the team of thirty-six scorers (secondary schools and universities) who evaluated the original 1600

Interface essays, and the almost two hundred teachers who participated in the Intermediate Evaluation Project supplying appropriate criteria and assisting in the holistic scoring last spring. The subsequent analytic scoring and error counts for a sample of the Intermediate essays were conducted by a team of four under Paul Legge, Consultant with the London and Middlesex Roman Catholic Separate School Board. The further analyses (syntactic maturity both Intermediate and Interface essays, error counts and analytic scoring for Interface essays only) was conducted by a group of eight experienced teachers in Sudbury as part of the contract which permitted this report.

The drawing of the sample for analysis, the supplying of technical information, and computer-based analysis are thanks to Dr. Ross Traub, Head of the Educational Evaluation Centre, and Mr. Richard Wolfe, Head of the Computer Services Group, at OISE.

At early stages of both the Interface and Intermediate Projects, Dr. James Britton was immensely helpful with advice and caution. This was during his stay as Visiting Professor at OISE.

Finally, many parts of the analysis and portions of the report on which this monograph is based are the work of Marion Marsh, Research Officer at OISE's Midnorthern Centre in Sudbury.

Peter J.A. Evans
Head, Midnorthern Field Centre, Sudbury
The Ontario Institute for Studies in Education

1 Purposes of the Study

In the past several years, there has been a great deal of interest on the part of the public and of professional educators in student performance in writing. It is generally alleged that standards have deteriorated, that students cannot write effectively, that today's graduate, if not illiterate, is at least ill-prepared for the demands that the world of work or of postsecondary education makes on the student's ability to express himself adequately in writing.

Most of the rhetoric, almost all of it negative, has been on the basis of little if any specific information showing exactly how students do write. It is simply a fact that, until the Interface Study in 1976 and the OISE Intermediate Writing Evaluation Project in 1977, no one in this province had set out systematically to gather or to analyse the writing of students either in elementary or secondary school programs.

At the same time, many teachers and others with responsibilities in our educational system believe that it is impossible to obtain reasonably reliable measures of student performance in writing and that effort to design an evaluation scheme for writing within a school or system, and certainly across systems, would be close to unmanageable.

This study has two main purposes:

1. *To describe in as useful a manner as possible the information we already have concerning student writing through the Interface Study and through the OISE Intermediate Writing Evaluation Project.*
2. *To explain how a school or a system can undertake evaluation of student writing performance and provide valid and reliable information to teachers and to the public about the state of writing at various grade levels in the school or system.*

Because full published reports are already available for the Interface Study, our description of the project will be confined to the writing sample and the scoring and analysis of the writing. The Intermediate Evaluation Project will be outlined in somewhat more detail in order to

provide the context and purpose of the assignments and to relate writing to program evaluation, one of the purposes of the Intermediate Project.

The various scoring procedures - holistic, analytic, error counts, and measures of syntactic maturity - will be described in sufficient detail for teachers to be able to extract and use those they consider most profitable. I shall try to avoid technical jargon as much as I can and to keep footnotes to a minimum.

The description of writing performance both at grades 12-13 and 7-8 will be given in much more detail, partly in response to requests for more readily available information on student writing and partly to illustrate the types of information various scoring procedures provide. Some of the counts we used turned out to be rather profitless; we want to share that information, too, so that readers may judge more effectively which strategies, given time and budget constraints, to select for their purposes.

The original report contains limited comparisons of grades 8, 12, and 13 writing. The present monograph will make these comparisons even more briefly and generally as observations, since conditions under which the two studies were conducted were quite different. These differences are explained at length in the report.

The original report offers some comparisons of writing performance of our grades 8 and 12 students and the writing of thirteen- and seventeen-year-olds in the American National Assessment, as well as some observations on writing development to be found in other research studies. For details, please see chapter 6 of that report.

The most important reason for evaluation, surely, is in order to make those changes likely to improve the quality of writing and of the writing/language program. At various points in the description chapters and in the concluding chapter I have made some tentative observations on reasonable program implications derived from our analyses. Here, as in the November 1977 report to teachers participating in the Intermediate Project, I wish to stress that the observations and speculations are indeed tentative. I have my own hunches and biasses. Teachers and consultants may in planning their English programs wish to act upon some of the data in the present study and to set out their own strategies to gather the best program-related information. What we all

certainly need is more evidence - a clean, hard-nosed examination of what students actually do in their writing - before riding off in one direction or another. I hope this report will stimulate educators to engage in regular, systematic evaluation of writing programs, lifting whatever they find useful from the methods employed here.

A Few Terms for Convenience

Terms such as "holistic scoring" and "analytic scoring" will occur frequently in the following chapters. At appropriate points these will be explained in some detail, but it may be helpful to the reader if they are briefly explained at the outset:

Holistic Scoring. Rapid impression scoring on an arbitrary scale. The marker reads rapidly a fairly large number of essays and ranks them on the scale in relation to each other rather than to external criteria. He makes no marks or comments on the paper. The marker is to use all points on the scale but is not required to produce equal piles or to make his distribution fit the "normal" curve. Nothing on the essay should identify the student in any way - grade, program, or school.

This strategy has the advantage of high reliability or consistency, much greater than traditional scoring. It is indeed quick. Its disadvantage is that it provides no information about the characteristics of the writing; its use, therefore, is specifically limited to ranking writing within a sizeable group or across groups or grades.

Holistic scoring was employed both in the Interface Study, and in the Intermediate Evaluation Project (for the final pair of assignments). A scale of 1-10 was used with 1-low and 10-high. Essays in the following chapters are identified by their rank order on this scale, and all bandings, quartile or percentile, are based on essay distribution by grade level on this scale.

Analytic Scoring. The scoring of essays, again using an arbitrary scale, on a number of clearly specified criteria.

In the Intermediate Project teachers developed a handbook of criteria appropriate to specific modes or types of writing. They were encouraged to use this handbook as a resource in scoring term assignments and in supplying information to students about their writing.

The technique was applied formally in the subsequent analysis of a

sample of the holistically scored writing last fall and informally to a sample of the grades 12-13 essays in the Interface Study proper. The technique was formally applied with a larger sample in the current study. A five-point scale was used for each criterion with two scorers and scores averaged. In the Interface Study modified versions of criteria elaborated by Paul Diederich (*Measuring Growth in English*, National Council of Teachers of English, Urbana, Illinois, 1974) were used.

Error Counts. Self-explanatory. In both studies a sample of essays was used for error counts. A large number of categories were employed. The counts were made independently by two scorers and the counts averaged.

Within the original Interface Study, the count was conducted on the first 200 words of the fifty essays scored by all scorers. In the present study the sample has been expanded to 200 essays and the count is made on the total essay. The Interface Study reports error frequency per 200 words; in the present study the frequency is per 100 words.

Measures of Syntactic Maturity. Syntactic maturity is difficult to define. The assumption is that as the writer matures, his writing style undergoes many changes. Thus, the effectiveness of the writer, that is, his ability to handle more complex syntax, to a degree can be measured.

Researchers have used a large number of strategies to make this measurement, and not always have they used the same strategy in exactly the same way. The most common is the length of the Minimum Terminable Unit (T-UNIT), which is defined as the main clause with all of its dependent material. Thus a T-unit and a sentence frequently coincide. Here are four examples illustrating our use of it in this study:

- He said that he would go.	1 T-UNIT
- She answered the phone and talked half the night.	1 T-UNIT
- She answered the phone/but she hung up quickly.	2 T-UNITS
- After he had hidden for a while, his mother called him/and he crawled out.	2 T-UNITS

Other measures, also used in this study, are sentence length, frequency of compound sentences, frequency of noun clauses and frequency with which a sentence opens with a phrase or subordinate adverb clause. Many more measures abound; we selected what we thought might be most useful in distinguishing weak writing from effective writing or in

distinguishing between the maturity of grade 8 and grade 12 writers. Not all of ours worked.

There will be numerous references to the Interface reports with which I was associated and to the report on grades 7-8 writing prepared last fall for the participating teachers. These I shall indicate in parenthesis as follows:

Peter Evans,
The Report on Grades 7-8 Writing November 1977, as (7-8 Report)

Ross Traub, Richard Wolfe, Carla Wolfe, Peter Evans, H. H. Russell, *Secondary-Postsecondary Interface Project II: Nature of Students*, Ontario Ministry of Education and Ministry of Colleges and Universities, 1976, 2 volumes, as (Traub V.1 or V.2)

H. H. Russell, Carla Wolfe, Peter Evans, Richard Wolfe, Ross Traub, Alan King, *Programs and Student Achievement at the Secondary-Postsecondary Interface: Interproject Analysis*, Ontario Ministry of Education and Ministry of Colleges and Universities, 1976, as (Russell)

Reference will also be made to the handbook of criteria developed within the Grades 7-8 Project. Like the above, its full title is a mouthful: *Criteria for the Evaluation of Student Writing Grades 7 and 8: A Handbook, OISE, 1977*. It will be referred to simply as "the Criteria Handbook."

2 The Evaluation of Writing in the Interface Study and in the OISE Intermediate Evaluation Project

The Framework of the Intermediate Evaluation Project

An informal history of the Intermediate Writing Evaluation Project appears in *Orbit* 42 (April 1978). In brief, the project began with a request from the North Shore Board of Education for some assistance in the design of a writing program across grades 6, 7, and 8 and the development of more effective evaluation strategies. Some of the materials used in introducing other schools and systems to the current project are the product of this initial work in one system: in particular, the identification of appropriate writing modes and specification of criteria appropriate to those modes. Analytic scoring, which will be illustrated in more detail in chapters 3 and 4, was also introduced for the evaluation of writing over the term to improve teacher←→student communication about the student's writing and to provide means for recording individual student progress. (More detail concerning the application of analytic scoring for these purposes is outlined in the 1977 *Intermediate English Guideline* of the Ontario Ministry of Education.)

Briefly, in context of OISE's Intermediate Evaluation Project under the general coordination of Dr. Mark Holmes, Field Development Coordinator, the writing evaluation project extended in the 1976-77 school year to over 6000 students in eighteen Ontario school systems. In the school year 1977-78 the project expanded to include twenty-nine school systems.

The purposes of the project were as follows:

Development of the Writing Program

1. Assistance in development of a clearly articulated writing program for the grade or grades concerned, with greater cooperation of teachers in the planning.

Formative Evaluation

2. Specification of criteria appropriate to the writing mode and the grade level. These must be clear to the student as well as the

teacher, hence a sound basis upon which the student can work towards improvement in writing.

3. Improvement of the means of recording progress through the writing folder and in communication of information to the student and parents.

4. Increased motivation for the student.

5. Improvement of the "audience" mode: The teacher is less the examiner, more the "coach."

Summative Evaluation: The Grade, School, or System Level

6. Development of a highly reliable (i.e. consistent) and valid method for evaluation of student writing, with built-in reduction of the unreliability that results from the differences of standards between teachers (interrater reliability) and differences in the student's performance owing to how he feels at a particular time or how he responds in a particular mode (writer variable), and differences of marker reaction owing to changes of topic and treatment (topic variable).

7. A rating system which can be used to assess the effectiveness of the writing program across the grades concerned and for identifying the individual student's rank in relation to all other students in the grade and sample population.

8. Provision of baseline information concerning the current state of writing, to be used as a basis for measuring improvement.

9. Detailed and specific information about the characteristics and quality of the writing at any grade level.

And as a result of two years' experience and with the help of the present report

Modification of the evaluation approach (both criterion-referenced scoring with improved student/teacher dialogue, and holistic scoring) for reasonably easy adoption and operation with a school system (any grades from 6-13).

Please note here the emphasis on development of a writing program and on formative (or ongoing) evaluation right in the classroom. These were and remained primary objectives and we are pleased to say, on the basis of our June 1977 survey of participating teachers, that these objectives were met in the project. Indeed, the objectives as outlined above are somewhat post hoc. They were newly elaborated following the survey of participants.

The present report, however, has chiefly to do with "summative" evaluation: the strategies for making from time to time a reliable and valid assessment of the state and characteristics of student writing. Without such data, claims concerning improvement or decline in standards of performance lack any solid evidence. It is apparent that in the present climate such data need to be gathered both to make the public aware of the actual state of affairs and to provide the basis for recommending modifications in program and assessing the impact of such modifications. That is to say, summative evaluation is both healthy and necessary for any school or system.

In alluding to the context of the grades 7-8 writing to be examined in this study, we do wish to stress that summative evaluation must be based fairly on the legitimate objectives and emphases in school programs. We claim in the Intermediate Writing Evaluation Project to have done that.

Early in the first year of the project (1976-77) we surveyed participating teachers to discover which modes of writing were considered most important in grades 7 and 8. There was a large degree of consensus on such modes as

- First-person narrative real or imagined
- Third-person narrative with description
- Exposition: process/procedure ("How to...")
- Exposition: point of view.

Several other modes or variations of modes also proved popular. Using these, we put together a package of assignments (topics suggested by participating teachers) for use over the school year. With these we provided directions for analytic scoring. The student's final writing score in the project consisted of 60% based on four of the term assignments analytically scored by the student's own teacher and 40% based on the final pair of assignments supplied by us and scored holistically.

Additionally, we surveyed participating teachers to identify specific criteria to be applied to grades 7 and 8 writing - both general criteria and criteria specific to the various modes. Following the survey we asked the participants to elaborate these criteria (two elaborated per school with duplications). These were edited into the Criteria Handbook which was then sent to all participating schools for use in the analytic scoring. It was from this handbook that we drew in specifying criteria for the analytic scoring of the sample drawn from the holistic scoring.

The Final Pair of Assignments

In mid-April 1977 all participating students wrote on consecutive days a composition on each of the following topics, the first being first-person narrative with emphasis on description and the second being process/procedure exposition. The assignments as given here are those provided to the students.

NARRATIVE

We have all seen a room in a mess - a bedroom, a playroom, a kitchen or a bathroom - whether because a young brother or sister got loose, or we did, or because of some other disaster such as a wide-open window during a storm.

Start with the mess, explain the cause, and develop a short, short story in which, perhaps with a little parental guidance, the mess is cleaned up and at last the hero or heroine of your little story comes out all right (escaping punishment, or getting to the circus or fishing trip or whatever after all, or learning his/her lesson).

Don't put more than three (perhaps four) characters into your story. One of the characters, if you wish, may be yourself.

The story should be about two pages in length. Give the story a title.

Pay attention in your description of the room to accurate detail, colourful words, and make sure events develop quickly to a satisfactory ending.

You might start with the sentence: "The bedroom (or playroom, etc.) was a mess." Of course, you don't have to.

EXPOSITION

Imagine that you are explaining to someone about your own age, someone who has had little or no experience with the subject, how he or she can become reasonably competent in a specific skill in some sport or game.

You are not to describe the game or its rules; just explain how a person should develop or practise a specific skill.

For example: batting a softball or baseball
the jump shot in basketball
the art of pegging in a cribbage hand
developing a golfing swing

Your composition should be about three paragraphs in length and informal in style. Give it a title.

A touch of humour is most desirable, but be sure by the end that the reader would be able to follow your instructions with good results.

In 1977-78 (the second year of the project) the expository mode was changed to "point of view."

These April essays were each holistically scored by two scorers (the teachers participating in the project) on a scale of 1-10, and their

scores averaged. Thus we have the averaged scores on each composition and a score for the student consisting of the sum of the averaged score on each piece of writing. This score was combined with the term mark with 60% contributed by the term mark. For the holistic scoring, identification numbers were used so that no scorer knew the student's name, school, or grade.

From this set of about 12 000 pieces of writing the sample for analysis was drawn. When the packages of essays were distributed for scoring, 10 Narratives and 10 Expositions were duplicated and buried in the sets so that they would be scored by all markers. These scores were used to calibrate the markers so that scores could be adjusted for scorers who tended to be fairly consistently harsh or gentle in their use of the scale. Additionally, we drew at random 100 more compositions in each mode and duplicated and buried these for scoring by at least eight of the scorers. For these 220 compositions, then, we have highly reliable scores, and these constitute the sample under analysis. In the following analysis these 220 (actually 218 as two did not photocopy well) are identified in rank order by mode, with distribution (percentile bands or quartiles) based upon the distribution of grade 8 scores.

In the sample, grades 7 and 8 compositions are mixed. However, because all compositions were treated without reference to grade level in the scoring process, we have not differentiated between them in this sample other than to reference their characteristics to the grade 8 rather than the grade 7 or the combined score distribution. Thus compositions representing, for example, the first quartile are compositions with scores which fall in the top 25% of the grade 8 scores; similarly compositions representing the 40-60th percentile band have scores that fall in that part of the range (better than 40% of the compositions and worse than 40% of the compositions in that mode) for grade 8 students. I should add that these band and quartile divisions are approximate as, at the time the original report (November 1977) was prepared, I did not have available the exact distribution. The Interface grades 12-13 distribution, on the other hand, is exact.

Several matters require stress before we move to the analytic report on the writing:

1. Control of the topic variable:

In any study comparing student writing performance, it is very important

that the topic and mode be kept under control. When a wide choice of topics is given, students respond in a great variety of ways, and it is virtually impossible to compare the writing of students responding to different stimuli. As well, different topics produce different reactions on the part of the readers or markers. For this reason, we confined each April assignment to one quite specific topic, although, as is evident from the instructions given to the student, the student had a good deal of freedom in his response to that topic. Our examination of the writing produced showed that students found sufficient freedom within the topic restriction to write very interesting compositions. We did not find that the topic restriction was harmful to students in putting forth their best effort.

2. Validity and the conditions under which the final pair of assignments were given:

The modes of writing selected for the final pair of assignments were two that were found, in surveying grades 7 and 8 teachers, to be types of writing normally expected at this level. Students and teachers received information well in advance of the assignment days as to exactly what would be expected in those assignments, so it was possible for the students to practise in those modes and for the teacher to review expectations. We did not want any surprises on the writing days except the actual topic. These modes of writing with sample topics had already been distributed to teachers at the beginning of the project as part of the package of possible assignments for the term essays. As well, the criteria appropriate for these were among those produced in the criteria handbook. Thus everything possible was done to see that the type of writing expected in the final pair of assignments fitted with important features of the grades 7 and 8 writing programs.

In our instructions to the teachers beforehand, we asked that dictionaries, thesauruses, and other resources be available to the students while they were writing this pair of assignments. As far as possible, we wanted to duplicate normal writing conditions for the student. In order to maintain security and fairness to all, we could not permit the student to take his piece of writing home for revision or to bounce his writing off classmates or teachers. However, we did attempt to allow adequate time for the student to write, revise, and recopy carefully. The allocation for Narrative was one hour and a quarter and for Exposition one hour. However, we asked teachers to allow extra time if students required it

for the recopying of the assignments.

We asked teachers at the time of giving the assignments to allow a few minutes for students to talk about the topic and to ask questions for clarification. We hoped this would provide a stimulus situation similar to that which would be given by the teacher on the occasion of a normal writing assignment. These are the instructions provided to the teacher for each of the assignments:

NARRATIVE

We hope the student will have little trouble getting the idea. A few moments for questions and discussion before the writing begins may help. Students may wonder "Which characters?" and the teacher might suggest three or four family members. Encourage students to treat the incident either humorously or seriously as they choose.

EXPOSITION

If students have questions about what constitutes a "skill," please clarify informally before they begin to write. A few minutes for informal exchange of ideas may help stimulate individuals to get started. Please stress that the "sport" or "game" doesn't have to be one requiring a lot of training. It might be simply a strategy for a card game or even hop scotch!

Evidently, our strategies worked out satisfactorily. Very few compositions come back to us incomplete. Even if the prescribed time was not quite adequate, it seems that teachers provided the additional flexibility necessary. The writing itself showed every evidence that students took the assignment seriously and were giving their best effort.

This feature is essential for any proper assessment of student performance in writing. In the Intermediate Evaluation Project, the final pair of assignments was set within the framework of a series of assignments over the school year, and the students were assured that they would receive their scores before the end of the school year. The rewards were direct; the project was not a research exercise solely for the benefit of someone else.

3. Adequacy of Sample of Student Writing:

As noted above, it is vital in any evaluation of student writing that the conditions of writing be valid. What the student is asked to do should represent fairly modes of writing which have received attention in his program, and the writing situation should be such that the

student's effort is perceived to count for him; his performance must matter to him. These conditions should be met whether one is only sampling performance at a grade level or whether one is, as we were doing in the grades 7-8 project, attempting to obtain a fair and reliable measurement of the individual student's performance.

Confidence in that measurement is increased by increasing the sampling of the individual student's performance. In the project, measurement of each student's achievement involved six pieces of writing: four over the term and two in the final pair of assignments.

Our effort to include this larger sampling and to be certain that each student provided us in the final assessment (holistic scoring) with two pieces of writing, on different days, in two important modes is strongly supported by research conducted in Great Britain in 1966 by James Britton, Nancy Martin, and Harold Rosen. Their report appears as Examinations Bulletin no. 12, *Multiple Marking of English Compositions: An Account of an Experiment* (Schools Council, HMSO, 1966): fortunately, it is still in print and well worth studying. Holistic scoring of essays in the British equivalent of the former Ontario grade 13 composition examination is compared with the regular treatment, with the holistic procedure proven far more reliable. Where, in the British examinations, students were required to write two shorter essays instead of one long one, reliability of scoring increased still more.

4. The Sample for Analysis:

Two hundred of the 218 compositions analysed were randomly drawn from the total set of compositions, and we are confident of the accuracy of their scores. We have a fair sample of the writing of students involved in the project. At the same time, it must be stressed that this is not a representative sample of the province, and it would be incorrect to generalize to provincial performance at grade 8. We can, however, with such a large sample, speak with some confidence about characteristics of the writing of 13-14-year-olds, and we believe our report will have interest and use to those wishing to establish reasonable criteria for performance at this level.

The Context of Writing Evaluation in the Interface Project

In 1976 a major study was undertaken under the direction of the Ontario Ministry of Education and Ministry of Colleges and Universities to study

expectations, programs, and student performance at the "interface" between secondary schools and university. With a project so large in scope, three teams were required to conduct the survey and analysis, and one volume in the report series (Russell) draws the findings on study of programs and student performance together.

An issue of great interest to all parties and to the public was performance in English by senior secondary school students, and under Dr. Ross Traub, principal investigator for Phase II, "Nature of Students," my chief responsibility was conduct of the evaluation of English achievement.

A good many complex issues emerged concerning appropriate ways of measuring English performance; these are dealt with at length in the report (Traub V.1 and V.2) and will not be repeated here. The project operated under a severe time restriction and it was not possible to develop a broad range of tests and testing strategies. We limited ourselves therefore to extant tests of reading comprehension and language, and to the evaluation of writing. Not all of the students tested in English wrote the essay; altogether about 6000 grades 12 and 13 students were tested for English competence, and of these 1579 wrote the essay (771 SSHGD and 808 SSGD - all programs).

There were some problems in sampling and a number of constraints imposed on the testing in the Interface Study. They are documented in the Traub report *Nature of Students*, vol. 1, pp. 16-20.

The Test Advisory Committee was fully agreed that, if only one writing assignment was possible, it should be an Exposition in which the student presented a point of view or argument. This choice of general mode was strongly confirmed in our following "Test Appraisal Inventory" by teachers at both secondary and postsecondary levels. The Test Advisory Committee, however, would not agree to provide only a single topic. Consequently a wide choice of topics was available to the students with consequent loss of reliability in the scoring.

The students were given one hour and a quarter to write a 250- to 350-word essay on one of eight topics. They were given the following instruction:

Develop your position with reference to the topic chosen, supporting it with carefully chosen illustrations and carrying it forward to a logical conclusion. Feel quite free to agree, disagree or take an intermediate position with respect to the point of view stated or implied by the topic.

These were the topics assigned:

1. *Should scenes of violence in books, in films, or on television be censored?*
2. *What are the values in the Olympic Games, not only for the competitors, but especially for the countries they represent and for the world?*
3. *"The progress of science guarantees the progress of civilization."*
4. *Secondary schools should not demand that students take particular courses or pre-requisites; students should be free to choose courses of interest to them.*
5. *"A competitive environment develops strength of character."*
6. *The seventies are the best time to be young in Canada.*
7. *"One of the characteristics of our time is our unwillingness to participate; more and more we are a nation of spectators."*
8. *Does art imitate life, or does life imitate art?*

Obviously, despite the Advisory Committee's desire to restrict the mode, some of these topics call for very different kinds of treatment from others. Compare, for example, topic 4 and topic 8. Differences in student response make it difficult to treat the holistic scores with the confidence one might wish, as scorers had to react to different kinds of writing.

In the time available for writing, it was evident that most students had little opportunity to revise what they had written. Certainly, few had any opportunity to copy their work over. This fact likely accounts in part for the relatively high frequency of routine mechanical errors in writing. Both the reduction of topic choice and provision of more time for writing and revision should be guaranteed in future studies of this sort.

The essays were holistically scored by a team of thirty-six scorers. Two-thirds of that team were experienced teachers of secondary school English, and one-third were postsecondary instructors, either university or community college. Each team consisted of two secondary school members and one university or community college member. Although this team did not meet together, they were given by mail a trial run on a sample of essays prior to their conducting of the official scoring.

From this set of approximately 1600 essays, 50 were chosen at random, subject to the constraint that they reflect the overall pattern of topic choice. Specifically, 10 of the selected essays dealt with the first topic and another 10 dealt with the fourth because of the popularity

of these topics. Five essays were selected from those written on each of the other six topics. A further constraint was that the set must be representative of the grade level and sex of the student authors. These 50 essays were photocopied and buried in the bundle of essays to be marked by each scorer. The scores given by the thirty-six markers on these 50 essays were used for the calibration of the markers in the use of the scale. Most of these markers were able to participate in a later rescoring of the 50 essays, so that we were able to take into account variation of scorer behaviour over time.

These 50 essays, for which we had highly reliable scores, were analysed on the basis of criteria (much as described earlier for grades 7 and 8) and counted for errors. The results are reported in detail in Russell.

The Interface Study reports performance by groups (those continuing to grade 13, those continuing to university, etc.); in our analysis for this study we have described in general characteristics of grade 12 and grade 13 writing without regard to program or postsecondary expectations, with our chief emphasis on grade 12 performance levels. Consequently most of the tables referred to later are based on grade 12 score distribution. We have included the 50 essays originally analysed and added another 150 essays randomly drawn. The 50 have highly reliable scores (calibrated averages of thirty-six scorers); the scores for the remainder are the average of three scores.

Two Reservations about the Interface Sample

a) The Topic Variable

Unlike the Intermediate Evaluation Project where the topic/mode variable was under control, the wide range of topics and potential treatment of them by students undermined the reliability of scoring in the Interface Study. It is important that scorers asked to compare the quality of writing, as holistic scoring demands, be looking generally at the same type of writing. Not only do scorers each respond differently to different topics, but also selection of topics by students is far from random.

Following the Interface Study, Richard Wolfe, Head of the Computer Services Group at OISE, made his own study of topic bias. He used scores on the reading comprehension and language test as a predictor. His finding was this: the topics chosen by the more able students gave

these students positive additional advantage in the holistic scoring while the choice of topic by less able students tended to disadvantage them further. This bias was especially acute for Topic #1 chosen by many less able students and for Topics #3 and #8 chosen generally by more able students. This finding reinforces our recommendation of strict mode and topic control in formal appraisal of student writing. Topic restriction is more rather than less fair.

b) Motivation

In the Intermediate Evaluation Project, as described earlier, it was possible to build the final pair of assignments into the writing and evaluation program for the school year: modes were carefully chosen to fit program and the student received his scores before the end of the school year. In the Interface Study where populations, not individuals, were the concern, this was not possible. Consequently, lack of motivation was a serious problem, reflected in the absentee rate for the occasion, in what the students told us bluntly, and in the writing itself. Many students did not take the assignment seriously.

It would be unfair, therefore, both to the students and to the educational system as a whole to claim that the writing under analysis typifies what students can do (or would do if properly motivated) in the Senior Division in our secondary schools. In the Interface report, I said that the most serious problem with most of the writing was not error frequency, but sheer dullness: a lack of personal commitment and interest:

> While many commentators on student writing at the interface have been greatly concerned about error frequency, the frequency of such apathetic writing perhaps should be a matter of even greater concern. Doubtless the test circumstances were a significant contributing factor in that they did little to stimulate writer commitment; nevertheless, the dryness of so many essays suggests that more might be done in schools to encourage commitment and vigour in writing. The student with a stake in his ideas and with a richer sense of audience is the one who is likely to write purposefully and effectively. (Traub, V.2, p.81)

In our further consideration of the grades 12 and 13 writing in chapter 4, these reservations concerning the topic variable and the influence of conditions under which the essays were written must be kept in mind for any fair appraisal.

The Analysis

The 200 Interface essays have been analysed afresh with respect to positive characteristics (criterion scoring), error counts, and several measures of syntactic maturity. The 220 Interface essays were analysed for positive characteristics and error counts as part of the Intermediate Evaluation Project funded by the Ontario Institute for Studies in Education and participating school systems. Those analyses were the basis for the October 1977 report to participating schools. Subsequently, with Ministry of Education funding for the present report, measures of syntactic maturity were also undertaken.

The following chapters, chapter 3 for the grade 8 writing and chapter 4 for the grades 12-13 writing, present the highlights of those analyses. I have not gone into great detail concerning measures of syntactic maturity for we have found, in analysing writing at a particular grade level, that these measures have limited value. They are more useful in comparing writing performance at different grade levels. Full details appear in the original report.

Chapter 5 of this monograph provides a few comparisons of writing at grade 8 and grade 12, both for their intrinsic interest and as stimulus for further study of the development of writing ability.

3 Performance: Grade 8 Writing

General Observations concerning Performance: Grades 7 and 8 Students

In the spring of 1977, when the final pair of assignments were given, a number of the schools participating were also involved in testing in other subject areas within the OISE Intermediate Evaluation Project. As well, some schools provided data on socio-economic status or IQ. It was interesting to find that the correlations with socio-economic status (father's occupation) and IQ were both quite modest. The correlation with IQ was 0.48 and with SES only .15, very close to chance. This low correlation suggests that students are far less disadvantaged in writing performance by their social background than most people believe. Another finding in the report, written by Dr. Mark Holmes, Coordinator of Field Development at OISE, was that girls' average performance in writing was better than boys'.

In the Writing Evaluation Project just under 2000 grade 7s and just over 4000 grade 8s were involved. While our attention in this report is focussed on grade 8 only, it may be of interest to learn that the mean student scores (i.e. scores on the two assignments averaged) for the April assignments brought to a 0-240 scale were 143 for grade 7; 155 for grade 8, indicating a very considerable difference between performance at the two grade levels. This evidence of difference is stronger in view of the fact that most of the grade 7s were from schools where grade 8s were also involved. This surely is good, even if predictable, news for teachers and administrators wilting under a barrage of criticism that schools do little to improve literacy.

The difference in performance for all students between exposition (process/procedure) and first-person narrative was also very considerable, especially across the mid-range. It seems clear on this evidence that students in both grades "carry off" first-person narrative more effectively than they do exposition in this particular mode.

Those who have read widely on the subject of student writing and language development are not likely to be surprised.* The claim there is that younger students are much more effective in writing with themselves as subject whether the experience is real or imaginary; they are "autobiographical" people. It is only as they grow older that they begin to deal effectively at a more abstract level and with subjects more detached from them.

Likely, as well, narrative of one sort and another occupies a large space in elementary school writing programs. That is a curriculum decision that should at least be reviewed by teachers of grades 7-8. While one may strongly agree that students should be encouraged to write in modes in which they are most comfortable and have the greatest likelihood of success, many school writing demands, both in and out of English class, call for the student to write in the "how to" mode quite frequently.

The length of the narratives considerably exceeded the length of the expositions. Average length for Narrative in the mid-score (40-60th percentile) range was about 330 words; for Exposition it was about 185 words. Generally, the higher scores were awarded to the longer compositions in each mode. Difference in length between the best narratives and those at the 15-25th percentile was not great (360 words as against 310) with a sharp decline for the bottom group to 190. The sharpest decline for exposition occurred between the top (95-100th percentile) group at 290 words and the 75-85th percentile group at 218 words.

Further details concerning length are provided on pp. 24-26 of the Ministry report and in particular in table 3.1 of that report.

[*Note: In this monograph I have reduced the number of tables and graphs but provided identification numbers identical to those in the original report. Here, for example, I have omitted table 3.1, and retained table 3.2.*]

Analysis of the Sample: (A) Procedure

When the essays were packaged for scoring, we drew 10 in each mode for scoring by all scorers. These were drawn more or less at random as long as they sampled poor, ordinary, and excellent writing. An additional 100 were drawn strictly at random from each mode for scoring by at least eight scorers. Altogether, then, 110 were drawn in each mode, and these

*See in particular James Moffet: *Teaching the Universe of Discourse*, Houghton Mifflin, 1967.

constituted the sample for analysis. Two of these did not photocopy well, so the actual sample consists of 109 essays in each mode. For these we have highly reliable scores, so the analysis is pegged in all of the tables against the grade 8 score distribution.

In our October 1977 report to participating teachers, we reported criteria performance and error counts for only the 0-5, 15-25, 40-60, 75-85, and 95-100 percentile bands based on scores. For the present study we have provided data on all bands highlighting in summary the first and fourth quartiles and the 40-60 (mid-range) percentile bands. This highlighting provides more compositions in each group and makes it easier to observe general differences in performance among "best," "average," and "weakest" compositions in each mode. On the other hand, important information is buried when essays are grouped, as will be apparent from examination of error frequency in the lower percentile bands for Narrative. The quartile figure suggests an extremely high error frequency across the quartile; in fact, the sudden leap in error frequency occurred about the 12th percentile, not the 25th. As well, differences from essay to essay were quite large. For example, in the twenty narratives that fell in the 40-60th band, the average frequency of the no sentence error was 0.23 per hundred words. However, 6 of the 15 errors occurred in one essay, 2 in another, 1 in seven of the essays, and none in eleven of the essays. It is obviously very dangerous to leap from a statistical generalization to the likelihood of the occurrence of the particular problem in the writing of a particular student.

Our analysis of syntactic maturity was confined to the first and fourth quartiles, and the 40-60 percentile band. These counts were largely experimental; we were attempting to determine which types of count would most likely provide teachers with information useful for writing programs.

As will be seen in the following tables, we also made a count of average paragraph length, though we are not sure that our findings are particularly useful. Obviously, most students are using paragraphs though the quality varies a good deal. In the American National Assessment in 1969, a good deal of attention was paid to the paragraph, but their findings were not particularly helpful to them as they found it difficult to provide adequate definitions for what constituted an "effective" or "satisfactory" paragraph.

Analysis: (B) Analytic Scoring for Positive Characteristics

For analysis of the narrative and expository writing we selected mode-specific criteria from the Criteria Handbook used in the Project, slightly adapting some to fit even more exactly with what the assignment called for.

The four criteria for Narrative:

1. *ORGANIZATION (Structure, Introduction, Conclusion)*

   High
   - *Setting and circumstances for story rapidly and sharply established so that story development begins early.*
   - *Conclusion is satisfying, completing the incident effectively.*
   - *Introduction stimulates reader's interest.*

   Medium
   - *Introduction takes rather long to develop in relation to total narrative, though complete enough/or/there is a lack of sufficient introductory detail on setting, circumstances to provide adequate background for the story's events.*
   - *Conclusion is either too abrupt, unexpected, leaving loose ends or trails off somewhat as if the student were not quite certain how to end, though end he must.*

   Low
   - *Lack of any effective introduction or conclusion.*
   - *The story begins without any detail which would capture reader interest or provide context. (Or student is so wound up in introduction and description that the story never properly gets under way.)*
   - *The conclusion makes little sense in relation to the story or as completion of the story.*

2. *POINT OF VIEW CONSISTENTLY MAINTAINED*

   High
   - *Narrator remains consistently in character, seeing events through his own eyes or those of his chosen narrator.*
   - *Language choice is consistently informal; style and phrasing are realistic and consistent with the personality of the narrator.*

   Medium
   - *As for "High," with possibly an infrequent lapse from first to third or from third to first person.*
   - *The writer/speaker's character is not always consistent in register or choice of language, but the over-all impression of viewpoint remains unified.*

   Low
   - *A consistent impression of the writer/speaker is lacking, so that a real and reasonably integrated picture fails to emerge.*
   - *There is inconsistency in the level of language employed.*
   - *The writer slides into the viewpoint of other characters or switches person.*

3. *LANGUAGE CHOICE*

   High
   - *Economy and terseness of language so that detail is brief but telling.*
   - *All details create a telling picture or contribute to atmosphere/mood.*

- *Language and detail are fairly consistently realistic.*
- *The language of the observer and the dialogue employed show a good ear for how people talk and feel (level, vocabulary, sentence length) in the situation being described.*

Medium
- *Tendency to use too many words in the attempt to convey the scene to the reader.*
- *Evidence of effort to convey the feelings, language of the people involved even though the "ear" for language in relation to the character speaking is not entirely consistent.*
- *Some excellent choices of words and images, though overall rather bland.*

Low
- *Language choice is woolly or trite. Description fails to convey any sharpness of detail or context.*
- *The language level and vocabulary show little evidence of adaptation to the characters, including the observer, in their situation.*

4. *SENTENCE STRUCTURE*

High
- *Evidence of variety (order, structure, length) related to purpose or effect.*
- *Some ear for sentence rhythm; occasional use of parallelism, antithesis.*
- *Inessential ideas are subordinated, and compound sentences are infrequent unless with a purpose.*

Medium
- *The student varies his sentences, though the variety may sometimes appear accidental, or simply for the sake of variety. He may tend to abuse various exclamatory devices (exclamation mark, dashes, rhetorical question) merely to "perk up" the writing.*
- *The student has not yet mastered effective subordination (through word, phrase, clause) so some sentences tend to ramble, or occasionally unrelated elements intrude.*

Low
- *Too many choppy sentences and/or rambling compound or compound-complex sentences tacked together with conjunctions.*
- *Any variety appears haphazard.*
- *The student has not developed the ability to subordinate less important ideas; consequently the reader has difficulty determining what is important.*

The four criteria for Exposition: process/procedure:

1. *ORGANIZATION (Sequence or Outline)*

High
- *Clear statement of topic or purpose.*
- *Completely clear sequence.*
- *Sufficient, but not excessive, detail (unless details and setting added for entertainment in a humorous presentation).*

Medium
- *Statement of topic and purpose is clear, as is general sequence, though details of sequence may be a little woolly. Nevertheless, the reader can follow the steps or stages without serious confusion.*
- *Illustration, where needed, may be lacking or insufficient.*

Low - *Sequence is confused whether by steps out of order, by excessive and irrelevant detail, or by omission of steps.*
- *The reader has to supply his own information in order to make sense of the procedure.*

2. *STYLE (Competence, Confidence, Interest)*

High - *The writer is assured, familiar with his subject, and at ease in desribing it.*
- *The writing may be flavoured with mild irony or humour.*
- *The writer by a variety of devices, especially in his tone in addressing the reader, reveals his interest in what he is describing; thus his enthusiasm is "catching."*

Medium - *The writer is reasonably assured, familiar with his subject, but through his style conveys little personal interest in what he is talking about and certainly arouses little.*
- *The "medium" essay is characterized by the writer's competence, but it is rather dull.*
- *An excessively breezy, casual style may also reduce the "high" essay to "medium." [But be generous; enthusiasm is precious and in small supply.]*

Low - *The writer is obviously not very interested and seeks to arouse little response. What he writes is a set of unwilling notes. As well, the writer appears to know little about the subject.*

3. *LANGUAGE CHOICE*

High - *The writer has clearly in mind a defined audience whether it is one relatively familiar with his subject or not, so that language choice - degree of technicality or precision - is consistently appropriate to that audience.*
- *All terms are sufficiently precise that the reader can picture the steps or process without ambiguity.*
- *Illustrations or imagery is similarly effective.*

Medium - *The writer sometimes loses sight of his audience so that his language at times may be either too technical or too vague.*
- *There may be occasional ambiguity or inadequate illustration.*

Low - *Word choice is so vague that the reader has great difficulty in following or picturing the process or procedure.*

4. *SENTENCE STRUCTURE:* *As for Narrative*

We asked, respecting language choice and sentence structure, that scorers as far as possible leave out of account actual errors, as we should be catching these in the error counts; the writing in the analytic scoring procedure was to be judged from the positive standpoint: the presence or absence of desired characteristics.

The scorers used a 1-5 scale (1, low; 5, high). When the narrative

scoring was complete, we found that scorers had a tendency to use 1 or 5 very seldom and when this tendency combined with the effect of averaging scores the discrimination scale was shrinking too much. We intervened, asking them to try to use the 1 and 5 more often for Exposition.

In future studies of this sort, I strongly recommend use of a 6-point scale rather than five in order to counterbalance the shrinkage resulting from the averaging of scores. (A 5- or 4-point scale is quite adequate for scoring individual assignments for the student.)

In their scoring, both analytic and error counts, team members were given the sets of compositions in a random order without the scores awarded in the original holistic scoring. For each pair of scorers, scorer #2 received his set in an order the reverse of that for scorer #1 so that the effect of tiredness or tedium would be averaged out.

The two scores on each criterion were averaged except where a pair of scorers happened to disagree by more than 2 points on two or more criteria. These were refereed by a third scorer. Scores on the four criteria were not summed.

The score summary, bands and quartiles, appears as Table 3.2. Graphs are provided in the Ministry report, pp. 34-37.

If we examine the mean scores for the first and fourth quartiles and for compositions in the 40-60 band we observe a positive correlation between holistic (total impression) score and scoring on all criteria, either mode. This, of course, is to be expected. The question is whether this type of analytic scoring merely duplicates, quite unnecessarily, holistic scoring or whether it reveals additional useful information.

On several occasions from band to band we note that the analytic score increases as the holistic score decreases, and for substantial parts of the range, while score continues downwards, there is little change in the average score on a particular criterion across several bands (e.g. Exposition: organization across 40-60, 25-40, and 15-25, about 45% of the compositions). For all criteria in both modes we find near the mid-range (Exposition) or just below it (Narrative) either a "flattening" or a slight upturn in quality before the plunge to the bottom. In all cases that plunge occurs in the 5-15 band for Exposition and in the 0-5 band for Narrative.

These "against the grain" performances of the analytic scores suggest

Table 3.2/ Grades 7-8 Positive Characterization (Grade 8 Distribution)

(1) Narrative

%ile Band	N	Organization	Point of View	Language Choice	Sentence Structure
Quartile 1	*34*	*3.59*	*3.74*	*3.68*	*3.50*
95-100	9	3.78	3.61	3.83	3.61
85-95	15	3.40	3.60	3.67	3.37
75-85	10	3.70	4.05	3.50	3.60
60-75	21	3.14	3.26	2.90	2.93
40-60	*20*	*2.73*	*3.18*	*3.13*	*2.90*
25-40	10	2.55	2.70	2.75	2.65
15-25	8	2.38	2.50	2.56	2.19
5-15	9	2.33	2.72	2.50	2.33
0-5	7	1.86	2.29	1.93	1.50
Quartile 4	*24*	*2.21*	*2.52*	*2.35*	*2.04*

(2) Exposition

%ile Band	N	Organization	Style	Language Choice	Sentence Structure
Quartile 1	*20*	*3.53*	*3.68*	*3.60*	*3.48*
95-100	5	3.70	3.80	3.50	3.70
85-95	5	3.80	4.10	4.00	3.80
75-85	10	3.30	3.40	3.45	3.20
60-75	16	2.94	3.31	3.03	3.03
40-60	*27*	*2.41*	*2.41*	*2.46*	*2.20*
25-40	14	2.57	2.82	2.61	2.54
15-25	9	2.56	2.44	2.61	2.39
5-15	15	1.70	2.27	1.93	2.13
0-5	8	1.56	1.88	1.75	1.94
Quartile 4	*32*	*1.91*	*2.22*	*2.08*	*2.16*

that there is some difference of importance between scores derived holistically and scores derived analytically and that the analytic scorers are having some success in discriminating one characteristic of the writing from another.

The absence of averages above 4.1 is likely less a reflection on general essay quality than it is on the narrowness of our scale and scorers' tendencies to avoid the extremes (1 and 5). The "bottom" of the scale we used was 1.0, not 0; there is no 0-1 interval. We should, I think, assume therefore that scores below 2.00 suggest unsatisfactory performance on the criterion. This occurs for three of the four narrative criteria in the lowest band (0-5) only, for two of the four expository criteria in the (5-15) band, and all four in the (0-5) band. In general, therefore, it seems reasonable to conclude that most students in grade 8 are performing adequately or well on a number of important criteria.

Each reader will wish to highlight for himself features of the criteria tables. I should like here to point out a few that seem to me to have some interest.

For Narrative, there is a sharp drop in the quality of organization below the 75th percentile, suggestive that most grade 8 students have difficulty in keeping the narrative under control, while ability to maintain a consistent viewpoint stands relatively well across the score range. A possible program implication is that teachers should give more attention to organization of Narrative: in practice this may mean asking the student to write shorter narratives limited to a specific situation or scene rather than to write loose and open-ended stories. The implication is not necessarily more practice in writing narrative but a fresh look at what is encouraged or expected when narrative is assigned.

The quality of organization across most of the score range for Exposition, and this was a limited and specific assignment, is a disappointment as well.

In general the quality of sentence structure fares less well than other characteristics of the writing. In the course of the Project, I read a large number of the compositions and would like to share an impression of the narrative writing with the reader. The problem isn't serious errors. It lies in the inability to define adequately units of thought, to subordinate less important material, and to build sentences for effect.

The sentences are very loose (in both the technical and non-technical meanings of "loose"), very close to the oral form of language: the breathless telling of incident crowded on incident without regard for effect beyond that breathlessness.

The reader may recall Dylan Thomas in *A Child's Christmas in Wales:* "I knocked my brother down, and he knocked me down; then we all had tea." Told aloud, it has its own brand of effectiveness; in print it is less effective. A good many of the students' sentences are like that.

At what point in the student's language/writing career do we apply the curbs to this kind of unselective vivacity? When do we say, "Hey, wait; you're writing now, not talking"? I shouldn't be in too much of a hurry. I'd like to encourage the vigour, not risk confining it too soon.

In the Narrative, with respect to language choice, many otherwise able writers are destroying their effect by sticking extra adjectives in everywhere. It seems likely that the student has discovered the teacher's reward system: the more colour words the better, relevant or not. Students substitute superfluity of detail for economy and effectiveness of detail. If teachers agree in this diagnosis, the means of improving the situation appears obvious.

Analysis: (C) Error Counts

The error counts were conducted separately from the positive characterization. While the same team was used, the error counter did not score the same essay for positive characteristics. Two scorers counted the errors in each composition and the results were averaged.

The narrative essays were counted first. In this exercise twenty-one subclasses of error were counted; for Exposition, the subclassifications were reduced to eleven, and these were the classifications later used for the grades 12-13 sample. Our scorers found the attempt to keep track of twenty-one categories rather burdensome and time-consuming; and, while frequency of errors in some of these classifications proved very interesting, the difficulties outweighed the benefits.

A further difficulty - even when classifications were simplified for grades 7-8 Exposition and Interface essays, and even with a practice run - was maintenance of consistency in classifying errors, especially in sentences where punctuation and sentence structure errors both appeared.

For minor punctuation errors we found a good deal of variation between scorers, chiefly because so many of the conventions of punctuation are undergoing change. A contributing cause no doubt was the faintness of some of the photocopies in which some marks that were small or faint even in the original were tending to disappear altogether.

The "rules," or at least rule books a generation ago, stressed frequent use of the comma. The tendency today even in quite prestigious quarters (I am told the University of Toronto Press for one) is to reduce minor punctuation to a minimum, the rule being to delete the comma following sentence adverbs and introductory phrases. Thus, in this decade, the sentence I am writing now would be more acceptable as "Thus in this decade the sentence..." The reader will observe the school of punctuation in which I was trained — eighteen bloody rules for the comma — and pray that I never get loose counting comma faults in his students' essays.

I am not being facetious. The rules for punctuation are much more fluid than any standard text concedes, and particularly subject to whims and trends. It would be unreasonable to expect complete agreement between any two error counters no matter how detailed the scoring rules might be. Readers may conclude that a count on punctuation errors, then, is not worth the trouble, and they may be right.

In the counts that follow, in almost every band punctuation "errors" account for a very high percentage of the total errors: for grade 8 Exposition the percentages of errors that were punctuation errors were, for Quartile 1, 55%; for percentile band 40-60, 58%; and, for Quartile 4, 50%. (And these do not include punctuation errors responsible for major sentence problems.)

In fairness to the students (7-8 or 12-13), please note that our error counters were instructed to count remorselessly for punctuation errors using pretty well the traditional rules. Many of the errors they counted more liberal markers would let go. This must be remembered in any fair evaluation of student performance from the standpoint of error counts.

All of the error count tables in chapter 3 and chapter 4 employ the classification code which appears in the left-hand margin of the classification. For those tables we combined the grade 8 Narrative counts as we have done for the others. In effect, then, with one exception (quotation

CLASSIFICATION AND CODE

		8 NARRATIVE	8 EXPOSITION	INTERFACE ESSAYS
(1) CONVENTIONS				
CODE				
Sp.		Spelling	Combined as Spelling	Spelling
Ap.		Apostrophe		Apostrophe
End-Pc		End punctuation	End punctuation	End punctuation
Pc		Punctuation	Punctuation	Combined as punctuation
	Q	Quotation marks	Not included	
C		Capitalization	Capitalization	Capitalization
(2) SENTENCE STRUCTURE				
Maj.		Major Errors in Sentence classified separately as	Major	Major
	Fr	Fragments		
	Fu	Fused		
	RO	Run On		
	C	Comma Fault		
Min.	K	Awkward Sentences (e.g. misplaced modifiers, faulty ellipsis, minor errors in structure)	Awkward	Awkward

		8 NARRATIVE	8 EXPOSITION	INTERFACE ESSAYS
(3) GRAMMAR				
Verb - Vb	S-V	Subject-Verb Agreement	Combined	Combined
	PP	Principal Parts (all spelling errors of principal parts included here)		
	T	Tense sequence		
Pronoun-Pro	PA	Agreement	Combined	Combined
	PS	Sequence		
	C	Case		
Other-Oth.	C/P	Conjunction/ Preposition	Combined	Combined
	OTH	All other Grammar errors		
(4) DICTION				
Diction	W	Wrong word, i.e. seriously redundant or faulty in meaning	Combined	Combined
	S	Slang or inappropriate register		

mark errors deleted in grade 8 Exposition) the counts are all based on the same general classifications and conducted under the same rules for both Intermediate and Senior writing.

We have noted the difficulty of classifying errors, especially faulty punctuation related to sentence errors. Another group difficult to classify is the inappropriate or unidiomatic use of the preposition (not an ungrammatical use). Many of the diction errors concerned the preposition, and in some marginal instances these may have been misclassified under "Other Grammar." The distinction between "Diction" and "Spelling" is difficult (indeed arbitrary) for homophones. We have arbitrarily assigned misspellings of principal parts of verbs to "Verbs" rather than to "Spelling."

In planning the error counts we attempted to be rigorous in order to provide as much information about the writing as possible. Some readers may feel that we have exceeded our mandate, but we hope that what the tables do reveal will be helpful to teachers in reviewing the kinds of emphasis to be given in language and writing programs.

A note on Narrative: Though we have not reported all of our subclassifications in the following tables, some readers may be interested in a more detailed breakdown of the data. The following are the subclassification figures for sentence structure and grammar for the mid-range (40-60th percentile) grade 8s:

Error frequency:

Sentence:	Fragments (no sentence)	- 1 in 400 words
	Fused sentence	- 1 in 500 words
	Run-on sentence	- 1 in 500 words
	Comma fault	- 1 in 300 words
Verbs:	Subject-Verb Agreement	- 1 in 1800 words
	Principal parts	- 1 in 700 words
	Tense/Tense Sequence	- 1 in 500 words
Pronouns:	Agreement, Sequence	- 1 in 1800 words
	Case combined	

We are looking then in the "average" range at a frequency of "no sentence" errors of fewer than 1 in 400 words and of total pronoun errors of 6 in 10 000 words. Subject-verb agreement errors are also 6 in 10 000 words. This appears to me a strikingly different situation from that claimed by those who tell us "Students can't write a sentence" or "Students can't write grammatically." They can and they do.

Conventions

In the following tables, errors are presented under two basic classifications: "Conventions and Everything Else"- Sentence, Grammar, Diction. All punctuation errors that assist in creation of a sentence error are classified under sentence errors, not with punctuation conventions. End punctuation errors ("End-Pc") refer to the failure to use a question mark rather than a period, or careless omission of a period (evidenced by the fact that the next sentence begins with a capital letter); they do not include "fused sentence" errors. We have made the distinction between Conventions (the courtesies of language) and Everything Else to make a point. Obviously convention errors outweigh all other classes; equally obviously, as we considered under our difficulties with punctuation error counts, many conventions are both arbitrary and liable to rapid change. Thus, when critics claim that students are "barely literate," it is salutary to note, however strongly you or I individually may feel about spelling or commas, that a large proportion of the errors are not in grammar or syntax but in less basic aspects of language. Such data in hand, we may be less likely to react inappropriately and negatively when faced with abusive criticism. Whether students know the objective predicate case or not, they write grammatically.

Tables 3.3 and 3.4 on the following pages lay out error frequencies per hundred words in percentile bands based on the grade 8 score distribution with subtotals for Conventions and Everything Else. Table 3.5 provides a summary of "top" (Quartile 1), middle range, and "bottom" (Quartile 4) in each mode and provides also proportions of the count (as percentages) for Spelling plus Apostrophe, Punctuation of the Sentence, and Grammar. Percentages of total convention errors and total "all other" are also given. Following the tabular information is a summary for each mode in graph form.

In the summary data (Table 3.5), convention errors, as already noted, are shown to outweigh all others. This is especially dramatic right across the range for Exposition where 4 out of every 5 errors are in conventions: Q1, 79%; 40-60 band, 81%; and Q4, 78%. For Narrative, in comparison, the split averages about 60% conventions, 40% "other" across the score range.

The relative constancy of error type by mode and consistent difference between modes suggests a close relationship between the type of writing

Table 3.3/Grades 7-8 Error Counts per hundred words (Grade 8 Distribution)

Narrative

Band	N	Aver. Words	Conventions					Sub Total	Sentence		Grammar			Diction	Sub Total	Total
			Sp	Ap	Pc-E	Pc	Cap		Maj	Min	Vb	Pro	Oth			
95-100	9	361	.23	.15	.05	1.09	.03	1.55	.23	.05	.23	.03	.10	.05	.69	2.24
85-95	15	401	.37	.10	.27	1.10	.07	1.91	.60	.12	.23	.05	.17	.08	1.25	3.16
75-85	10	349	.60	.14	.11	1.92	.20	2.97	.54	.29	.49	.09	.17	.17	1.75	4.72
60-75	21	389	.91	.29	.21	1.77	.17	3.35	1.09	.16	.54	.12	.27	.27	2.45	5.80
40-60	20	328	.84	.15	.63	1.25	.12	2.99	.96	.21	.40	.05	.29	.14	2.05	5.04
25-40	10	307	.68	.07	.26	1.92	.13	3.06	.95	.23	.59	.06	.33	.26	2.42	5.48
15-25	8	313	1.08	.20	.52	1.60	.52	3.92	1.84	.16	.68	.08	.20	.24	3.20	7.12
5-15	9	236	4.61	.33	1.65	2.64	.75	9.98	2.31	.14	1.65	.05	.33	.19	4.67	14.65
0-5	7	189	5.97	.30	.91	2.65	.68	10.51	2.12	.23	1.66	.45	.53	.45	5.44	15.95

Notes: 1. Spelling and Apostrophe errors were counted separately in Narrative and together in Exposition.

2. Errors in the use of quotation marks were counted under "Punctuation" for Narrative and not counted at all for Exposition.

Table 3.4/Grades 7-8 Error Counts per hundred words (Grade 8 Distribution)

Exposition

Band	N	Aver. Words	Conventions				Sub Total	Sentence		Grammar			Diction	Sub Total	Total
			Sp(Ap)	Pc-E	Pc	Cap		Maj	Min	Vb	Pro	Oth			
95-100	5	291	1.30	.62	2.20	.14	4.26	.21	.21	.14	0	.14	0	.70	4.96
85-95	5	265	.75	.23	1.88	.08	2.94	.45	.08	.15	.08	0	0	.76	3.70
75-85	10	219	.69	.46	1.51	.09	2.75	.37	.14	.18	.09	.05	.18	1.01	3.76
60-75	16	235	1.01	.48	2.05	.08	3.62	.40	.21	.13	0	.08	.11	.93	4.55
40-60	27	187	1.07	.51	2.45	.10	4.13	.42	.10	.16	.06	.06	.08	.94	5.07
25-40	14	193	1.41	.92	2.11	.07	4.51	.78	.11	.11	0	.18	.07	1.25	5.76
15-25	9	169**	1.76	.85	2.44	.40	5.45	.45	.23	.23	.06	.17	.28	1.42	6.87
5-15	15	136	2.40*	1.42	4.31*	.25	8.38	1.13	.30	.54	.05	.20	.30	2.52	10.90
0-5	8	141	2.04	.62	2.49	.27	5.42	.53	.27	.27	0	.18	.18	1.43	6.85

* One essay in this band contained a large proportion of the spelling and punctuation errors, 14 and 30 respectively.

** One essay of unusual length (411 words) is excluded from this average.

Table 3.5/Grades 7-8 Error Counts (Grades 7-8 Distribution)

(1) Narrative

Band	N	Aver. Words	Conventions					Sub Total	Sentence		Grammar			Diction	Sub Total	Total
			Sp	Ap	Pc-E	Pc	Cap		Maj	Min	Vb	Pro	Oth			
Quartile 1	34	396	.39	.13	.16	1.31	.09	2.08	.48	.14	.30	.05	.15	.10	1.22	3.30
			16%		45%			63%	19%		15%				37%	
40-60 %ile	20	328	.84	.15	.63	1.25	.12	2.99	.96	.21	.40	.05	.29	.14	2.05	5.04
			20%		37%			59%	23%		15%				41%	
Quartile 4	24	248	3.43	.27	1.01	2.20	.66	7.57	2.07	.17	1.24	.15	.32	.27	4.22	11.79
			31%		27%			64%	19%		15%				36%	

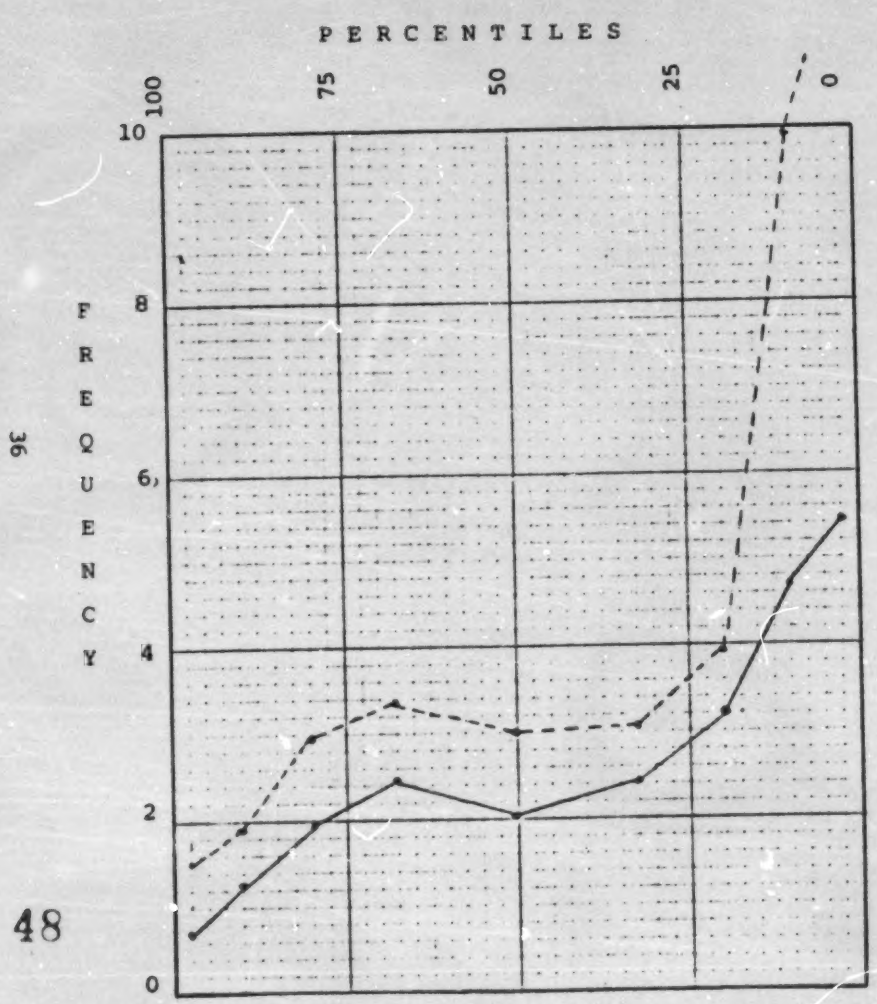
(2) Exposition

			Sp(Ap)													
Quartile 1	20	249	.88		.44	1.81	.10	3.23	.34	.14	.16	.06	.06	.08	.84	4.07
			22%		55%			79%	12%		7%				21%	
40-60 %ile	27	187	1.07		.51	2.45	.10	4.13	.42	.10	.16	.10	.08	.08	.94	5.07
			21%		58%			81%	10%		7%				19%	
Quartile 4	32	154	2.09		1.04	3.23	.30	6.66	.75	.26	.37	.04	.18	.26	1.86	8.52
			25%		50%			78%	12%		7%				22%	

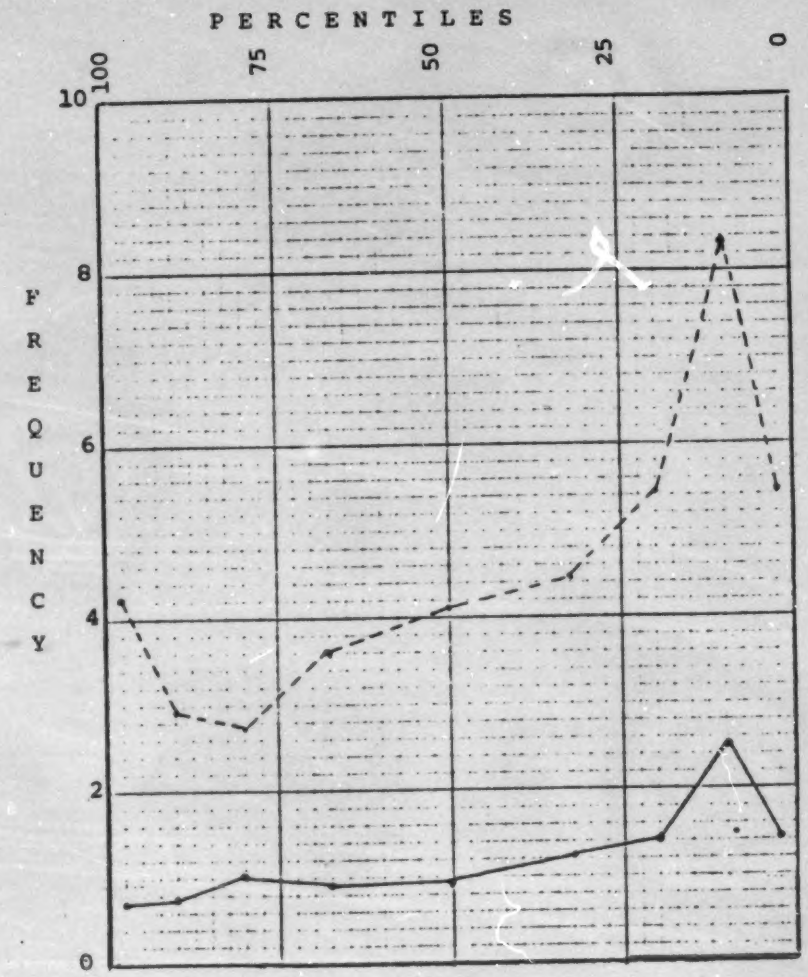
Notes: 1. For Narrative, errors in the use of quotation marks are included in Punctuation; in Exposition (because of low frequency of occurrence of quotations), they are not.
2. In Exposition, errors in the use of the apostrophe were included within "Spelling."
3. When scores of the two error counters were averaged producing frequently 0.5 fractions, 0.5 was treated as 1, 1.5 as 1, 2.5 as 3, 3.5 as 3, and so on.

ERROR COUNTS, GRADE 8

(1) NARRATIVE



(2) EXPOSITION



compelled from the students and, at least collectively, their performance with respect to errors. Though Exposition (process/procedure) is the mode in which students were less successful, this lower quality is evidently not the result of greater frequency of error in those categories, grammar and sentence structure, where error likely ought to be most influential on score. On the contrary, not only is the percentage of errors falling into these classes very much lower for Exposition, but the actual frequency of error per hundred words is often dramatically lower as well.

(From Table 3.5)

		Sentence Errors		Grammar Errors	
		Narr.	Exp.	Narr.	Exp.
Errors per 100 words	Q_1	.62	.48	.45	.28
	40-60 %ile	1.17	.52	.74	.34
	Q_4	2.24	1.01	1.71	.59

I have a hunch about this phenomenon, a pattern that contradicts expectations. My hunch applies particularly to the high frequency of sentence errors in Narrative, and it concerns the problem of sentence control previously discussed in Part 3 of this chapter. Might it not be that the average and weaker writers, caught up in some of the breathlessness of Narrative, simply don't yet have the distance on their writing to tighten sentences for best effect, the result being that as the sentence rolls on the errors almost inevitably slip in? In Exposition, on the other hand, with shorter compositions in the first place and a limited subject to discuss in the second, the student is not getting as carried away; more control remains.

It is unfortunate that none of our grades 12-13 writing was in the narrative mode. If it had been, we could compare frequencies of these problems at the two age levels to determine whether the high frequency, especially of sentence errors, in grade 8 should be a source of anxiety or whether many of these errors would disappear as the student matured. Here is a strong case, for schools or systems wishing to compare writing performance across several grade levels, to keep topic and mode under strict control for the assessment.

With the data we have, I cannot venture to turn my hunch into a hypothesis. But even with only these data we must conclude that counts of error frequencies taken in one mode only are insufficient to make generalizations

about problems pervasive in student writing. There is a very substantial difference for the average writer (40-60 band) in saying total sentence error frequency is 1.17 per hundred words and .74 per hundred words!

The summary of data in Table 3.5 is a summary of summaries (Tables 3.3 and 3.4). When we turn to those we can see how any abridgment of data buries important information. According to Table 3.5, the average total error count for Narrative in the fourth quartile is an alarming 11.8 per hundred words. When, however, we break our data down into percentile bands (Table 3.3) we observe that the dramatic jump occurs below the 15th percentile. Here, at perhaps the 10th or 12th percentile we arrive at a performance level at which we can distinguish "minimum competence" or "basic literacy" for the grade level.

And even Tables 3.3 and 3.4, as illustrated by a footnote, do not properly represent individual performance on any count. Often a preponderance of the errors in the essays examined for any band may appear in only two or three of the essays. While the kind of summative information gathered here may have value in making decisions about the writing and language programs in a school or system, the individual student remains cantankerously individual. How he writes can be determined only by close analysis of a good sample of his writing.

Obviously, and predictably, there is a negative correlation between scores and error frequency: as errors go up, scores go down. But, as we have already seen in comparing error frequency and scores between Narrative and Exposition, infrequency of error alone does not distinguish good from poor writing. When we examine on a graph Narratives and Expositions with respect to error frequency and score we observe across a very wide "middle band" little increase in error frequency, with the dramatic jump in frequency occurring only near the bottom of the score range, and there only for Narrative. Particularly startling is the remarkably "steady state" of error frequency in Exposition from the best to the worst writing; evidently error frequency here is not a very significant influence on writing quality.

Some error counts are so low as to suggest that the effort expended in compiling them is not worth the trouble.

I suggest in conducting error counts that a team (school or system) focus on a few targets, those that teachers consider most important from the

standpoint of their expectations for student writing for the age and grade. What is counted should count: that is, select areas of specific concern, areas in which, given some data, you may wish to make specific changes in program or to confirm the quality and utility of the present program.

Analysis: (D) Syntactic Maturity

Our purposes in attempting counts on features of sentences were to determine whether

1. some of these indicated changes in "syntactic maturity" between grade 8 and grade 12
2. some of these indicated differences in writing quality across the score range either in grade 8 or grade 12
3. some of these indicated differences in writing characteristics for different modes (narrative and expository) at grade 8

Our success was mixed, with greater success in comparison of students several grade levels apart, as comparison of the following Table 3.6 with the grade 12 (Table 4.6 in the next chapter) reveals. The T-unit measure, for example, appeared to be of little value to a teacher investigating the writing characteristics of students all at the same grade level.

An interesting review of some syntactic maturity measures has recently appeared (its arrival, unfortunately, too late to assist us) in context of recent studies of the writing of secondary and postsecondary students in New Brunswick.* The review suggests (as we found ourselves) that average sentence length was of little use in comparing writing at different grade levels. T-unit length, words per clause, and clauses per T-unit were more valuable indicators of maturation in writing.

The T-unit, which we did measure, is defined in Stewart as "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it" (p.37).

The counts we made in our analysis of grade 8 and grade 12 writing were

- average sentence length
- average T-unit length
- average frequency of noun clauses per essay

*M.F. Stewart, "Syntactic Maturity from High School to University: A First Look," in *Research in the Teaching of English*, vol. 12, no.1, Feb. 1978, pp. 37-46.

- compound sentences used (as a percentage of total sentences)
- sentences introduced (a) by a phrase (prepositional, infinitive, participial) and (b) by a subordinate adverb clause, expressed as a percentage of total sentences, plus their sum as a percentage.

Our scorers also kept count of the frequency of the noun clause used as subject, but its occurrence was so infrequent in both grade 8 and grade 12 writing that we have not cluttered the tables with the information.

Our scorers also kept track of paragraphs used and mean paragraph length. As noted at the outset of the chapter, this information has not proved particularly useful. It is clear that almost all students do use paragraphs with reasonable frequency, but this says nothing of their quality. Single-paragraph compositions were eliminated from our counts.

The realities about paragraph writing and the typical textbook theory and exercise materials ("Begin with a clear topic sentence," etc.) are in contradiction. In a study of paragraph development in such magazines as the *Saturday Review*, *Harper's*, and *Atlantic Monthly*, Richard Braddock made some fascinating discoveries.* First he had difficulty in practice in determining what a "simple topic sentence" or even a "topic sentence" was, and found instead that in a high proportion (45%) of the paragraphs examined, the topic was implied rather than stated. From his findings he also inferred that only about 13% of the expository paragraphs of contemporary professional writers begin with a topic sentence. He concludes:

> Teachers and textbook writers should exercise caution in making statements about the frequency with which contemporary professional writers use simple or even explicit topic sentences in expository paragraphs. It is abundantly clear that students should not be told that professional writers usually begin their paragraphs with topic sentences... (p. 301)

Neither should teachers. And, in light of this finding, it is not surprising that anyone attempting a handy set of criteria for effective paragraphing is going to have a great deal of difficulty.

The count we found to be of most value in comparing sentence style across a grade level was the frequency with which the sentence began with a phrase or adverb clause. This count is quite straightforward for the teacher to make quickly. There are other ways of "counting" to measure maturity in flexibility of sentence construction: frequency of participles and infinitives, frequency of sentence adverbs, frequency of

*R. Braddock, "The Frequency and Placement of Topic Sentences in Expository Prose," *Research in the Teaching of English*, vol. 8, no. 3, Winter 1974, pp. 287-302.

adjective clauses. These and other variations have been used in research studies. The choice of measures in any school or system study of writing should certainly include one or more measures of this sort, as results will have rather direct implications for teaching strategies related to sentence building.

The headings for Table 3.6 are self-explanatory with the exception of the T-unit, which is defined with illustrations. Compound sentences in which the subject was not repeated in the second clause were treated as one T-unit.

Evidently, on these data, the average sentence length provides little indication of differences across the score range or, surprisingly, between modes. The differences perhaps worth noting appear in compositions in the lowest bands. Interestingly, we find that the weakest narratives have by far the longest sentences (and T-units). As the number of compositions in that band is very small, we cannot consider the count especially significant, though it does tend to confirm my earlier hunch concerning the "out of control" narrative sentence. As well, that count should serve as a warning that T-unit length on a single or only a few pieces of writing is not a useful measure of syntactic maturity.

Paragraph length and frequency data appear not to be helpful in revealing differences across the score range in either mode, nor does the frequency (or, rather, infrequency) of the noun clause. Its infrequency in both modes suggests that most grade 8 writers do not have it in their repertoire. Incidentally, we specifically excluded from the noun clause count quoted speech.

I had expected to see a difference in the frequency of use of the compound sentence between Narrative and Exposition with a higher frequency for Narrative owing to students' tendency to tag actions together. Evidently this difference did not develop. An informal inspection of the process/procedure writing turned up a frequent use of the construction "First you do X, and then you do Y"; this construction likely accounts for the similar counts between modes. It will be interesting to see whether in the 1978 study, where the mode is "Exposition: point of view," we obtain different frequencies between Narrative and Exposition.

For Narrative, as in the Interface essays to be discussed in the next chapter, the frequency of sentences introduced by a clause or phrase

Table 3.6/Syntactic Maturity
(1) 7-8 Narrative

%ile Band	N	Sentence Average Length	T-Units Average Length	% Compound	Noun Cl. per Essay	Sentence Introduced by			Aver. No. Para	Aver. Length Para (words)
						Phrase As	Clause %	Total		
95-100	9	17.0	13.2	14.1	1.9	8.3	8.2	16.6	5.3	66
85-95	15	14.9	11.6	20.8	3.7	7.5	8.0	15.5	5.9*	87
75-85	10	15.8	13.2	13.1	2.6	12.8	12.2	25.0	5.1*	84
40-60	20	16.2	12.5	14.7	0.7	11.2	2.6	13.8	7.8*	37**
15-25	8	17.9	12.0	12.4	1.0	4.9	4.9	9.8	5.0*	very irregular
5-15	8****	13.4	10.3	26.1	2.5	0.6	15.8	16.4	4.6*	63
0-5	4	22.1	16.7	12.5	1.5	9.0	1.8	10.8	insufficient data	
(2) 7-8 Exposition										
95-100	5	16.7	13.7	15.2	0.4	21.8	11.4	33.2	3.8	81
85-95	5	17.6	12.2	10.1	1.4	3.4	19.6	23.0	3.6	92
75-85	10	19.5	14.8	23.0	0.6	14.0	11.9	25.9	4.1*	57
40-60	27	16.6	13.1	12.4	0.7	10.2	15.8	26.9	3.5*	57
15-25	9	17.5	13.8	8.8	1.3	4.1	27.2	31.3	4.2*	57
5-15	15	13.0	10.7	10.0	0.7	4.8	23.5	28.3	3.9*	42
0-5	8	15.8***	13.7	16.1	0.5	4.4	14.0	18.4	4.8*	40

*Several essays were not paragraphed. These are omitted from the count and length figures.

**An essay of 2 paragraphs averaging 256 words per paragraph is omitted.

Note: Figures in this table are averages of averages per composition.

*** One essay here consisted of one sentence of 63 words; it is omitted from the sentence, compound, sentence introduction, and paragraph counts.

**** One essay in this band conducted the story through dialogue fragments. It is not included in any of the counts.

seems to be of some value in distinguishing between more and less effective writing, though the frequency of 16.4% (almost all of it clauses) in the 5-15 band is surprising. The process/procedure writing, however, has a higher frequency through all bands reported except the last. In either mode the introduction of a sentence by a phrase is rare in compositions in the fourth quartile, suggestive, perhaps, that efforts to increase variety in sentence introduction might be well directed.

The data on average paragraph length are not helpful. We do note very short paragraphs in the weaker expositions, though this is more likely to be an effect rather than a cause of the underdevelopment these compositions tended to exhibit.

In both modes, although more frequently in Narrative, some students used a great deal of dialogue. Dialogue with its fragmentation of sentences and frequent paragraphing produces fits and starts in our counts from composition to composition. The T-unit length as a general indicator of syntactic maturity should likely not be used for counts in writing where dialogue predominates.

Compiling Information on Student Writing

Though our sample of grades 7-8 writing is not a provincial sample and is restricted to only two modes, the data summarized in this chapter provide much information about the general characteristics, strengths, and weaknesses in the writing of grade 8 students in Ontario.

This study and analysis likely includes much more information than individual teachers or even systems may wish to gather concerning writing performance. Further, any compilation of information should be carefully related to the purposes for which the information is intended: the desire to compare writing performance across one or more grade levels calls for different measures than those characterizing writing at a particular grade level. Measures should be related to program assessment - to the question of what might or should be done to improve writing performance at a particular level.

Some counts, notably sentence structure (positive), types of sentence structure errors, and categories of grammar errors provide very important information for language program planning. The surprising infrequency of some "traditional" error categories certainly has implications and is good news about young people and their writing.

For analysis of performance at any grade level, it is not necessary to analyse every essay. For a large system a 10% sample randomly selected (or randomly selected within programs) should suffice provided the mode/topic and conditions under which the assignment is given are controlled.

The analyses that appear most directly profitable are the positive characterizations on several criteria considered important for that mode of writing, but inclusive of organization and sentence structure, error counts in grammar and sentence structure, and some measure (e.g. the frequency with which a phrase or adverb clause begins a sentence) of sentence variety. Teachers in a particular school or system may wish to add other specific measures reflective of language concerns emphasized in the program. The measures taken here are quite manageable by teachers; for most, provided there is close prior understanding of the categories, a single scorer will suffice. For the criteria, however, average two scorers' evaluations and referee major differences.

SAMPLE INTERMEDIATE COMPOSITIONS

The following compositions provide a small sample of writing in each mode tested in the 1976-77 Intermediate Evaluation Project. The score range (holistic scoring) was 1.0-10.0.

The Narratives

1. Score 7.2, from the 95-100th percentile band:

All Because of a Box of Cereal

"Good Gravy, what happened here?"

It seemed like we were goners, forever through when mother walked in and discovered the bedroom where my kid brother and I slept. Earlier that day my brother Joe and I had gotten in a big fight over who could look for the prize in the cereal box. Before it had gotten to the point of one of us backing down, my mother sent us up to our room, to fight if we wished.

Who wants to fight when they are being punished, so we decided to play with some old Build-O-Blocks we had in our closet. You can imagine, a set of building blocks in the hands of two destructive boys. After less than an hour of constructing and knocking down our room looked like the Miami Dolphins had practice in it. Lamps crumpled on the floor, drapes hanging in strands, beds stripped of their covers and just about everything else close to extinction.

My mother has got to be one of the kindest people I know, so after an hour she came up to our room bearing milk and cookies and bringing the news of our permission to leave our room. Then she saw the room!

"Get this mess cleaned up immediately!" and she left. The thought of cleaning up your room isn't exactly pleasant so we decided to make a game out of it, basket ball.

First I toss a Build-O-Block into the block box, then Joe, then me, then Joe and so on till we had all the lamps, curtains, beds, and floor looking brand new.

Usually when my dad punishes us he doesn't spank or anything, he just deprived us of something we really want, this way of punishment hurts more I'm sure. Mom had told Joe and I that we could go to the fair so we were a little scared. Since dad was on one of his business trips though we might get away with just a severe waring.

Just our luck, Dad arrived home while we were still in our rooms building and wrecking. I can hear him on the stairs now. Have mercy!

I guess he had been expecting a garbage dump and instead he got a parlour. Dad was so shocked (our room is never clean) that he just went away laughing, I think that means we can still go to the fair!

2. Score 5.8, from the 60-75th percentile band:

The Family Room

My dad and I had just returned from our fishing trip and put the boat away. As I walked into the house everything seemed normal until I saw the family room. I couldn't believe my eyes. I walked into the room or attempted to I saw where the burgular had broke in. The sliding door had been opened with an axe and then smashed. All the other windous had been broken along with the lights. The television had been kicked in and the parts ripped out. The stereo was in pieces all over the room. He must have got something to drink because there were stains on the rug. The furniture was much the same as the rest of the room. The chairs were broken and had been burnt in the fireplace and the couch was upside down. The pillows and cussions had been ripped because there were feathers and spunge all over the room. The worst thing of all was the wall. The panelling had been cut and the wall partly destroyed.

Just then I noticed my neighbour peeking through the fence and it looked like he had a smile on his face. Could it have been my ex friend Bob Freeland? Just then my dad saw the family room and almost hit the ceiling. After he calmed down I told him about Bob. He said he was going to have a talk with Bob. Father asked Bob where he was last night and today and if he had seen anything. Bob couldn't take the pressure and finally admitted to it. Father had a talk with Bob's father and Bob's father agreed to pay for the damage.

After this incident took place I had a talk with Bob. He said he did it because I didn't give him rights to the pool.

3. Score 5.4, from the 40-60th percentile band:

The Bathroom

It was seven o'clock and my brother was told to take a bath by my Mother. My brother, if you did not know does not particularly like to

take a bath, especially when he is ordered to. My Mother finally talked him into it and he took a bath. My Mother and Father were going out that evening and I was left to babysit my antagonistical little brother at a mere fifty cents an hour. Well, incase you did not know my brother does not like to be ordered to do anything, so we just stayed out of each others way and minded our own business. I went in to see if my brother had cleaned the bathroom and sure enough my brother had left the bathroom in a violent frenzy. I called for him to clean it up and the little ruffian gave me a swift blow to the shins. I told him I would be forced to send him to bed early for his abrupt actions so he obeyed my orders and began cleaning the bathroom. Well, I went in to watch television for fifteen or twenty minutes. Figuring he was done by this time I went in and sure enough it was left just as messy as when he started. I called for him and found out he was over at the neighbours with a friend so I brough him back and started him cleaning the bathroom again. But this time I kept an eye on him. By this time it was nine-thirty and my Mother and Father would be getting home soon so I helped my brother with the task and sent him up to bed. Just then my Mother and Father walked in and said, "was he good." And I replied, "of course, he couldn't have been better."

The Expositions

1. Score 7.4, from the 95-100th percentile band:

Mounting a Horse

Proper riding begins with proper mounting. First get your horse under control by adjusting the reins evenly with enough tension to feel the bit and to hold the horse steady. But what ever you do, don't get the reins too tight because the horse may end up going the wrong direction. Hold the reins in your left hand and place this hand on the neck in front of the withers. Grasp the ridge of the neck or a lock of mane.

Twist the closest stirrup with your right hand and place your left foot securly in the stirrup. Brace your left knee against the horse and move your right hand to grasp the horn. You should now be braced against the horse with your left leg and hands making a triangle for support.

In good horsemanship there are two basic bor[illegible]positions. Mounting a strange horse could be difficult. Stand by [illegible]t sholder and be in position to see both ends of the horse. Make sure your infront of the saddle so if he moves you'll be able to swing into the saddle. If he does move, slap him, you may prevent a runaway. The second way should be used only on a calm horse. Stand near the stirrup and face towards the seat of the saddle, and then place your left foot in sturrip. In both cases spring up in the saddle and use both legs as stabiliizers. On the other hand if you don't like mounting horse's you could mount butterflies or stamps instead.

2. Score 5.1, from the 40-60th percentile band:

How to Saddle and Ride a Horse

Saddling up a horse is easy if you like horses, but if the horse doesn't like you don't plan on saddling him. A horse is a very sensitive

animal and hates people that hate horses. Saddling a horse takes about 5.00 minutes or less. When putting on a saddle on a horse it does not won't to suck in it's stomach so you have to hit it's stomach but not very hard.

First you have to put the blanket on the horse, then the saddle, do up the cinch, and then the strap with a buckle on it. After this is done put the bridle on and then get on your horse and put yourfeet in the stirrups and adjust to the size.

When mounting a horse you mount from the left side after you are on put your feet into the stirrups and leave some slack in the reins, kick the sides of the horse and ride off.

3. Score 4.8, from the 40-60th percentile band:

The Jump Shot

The first thing you have to do is get in shape. You should practise running, after you have your leg muscles in good shape for running you should practise jumping. After you have mastered jumping you should get your wrist muscles in shape.

To do this you should practise dribbling a ball or coconut or what have you. After you have mastered the running, jumping and have your wrist muscles in top shape you should try the jump shot. You should try the jump shot in the privacy of your back yard or some other private place.

After you think you have mastered the jump shot try it with a couple of friends. Now you are ready to try the jump shot in public. That is all you have to know except to always play fair. Good luck and have fun.

4 Performance: Grades 12 and 13

Context and Data

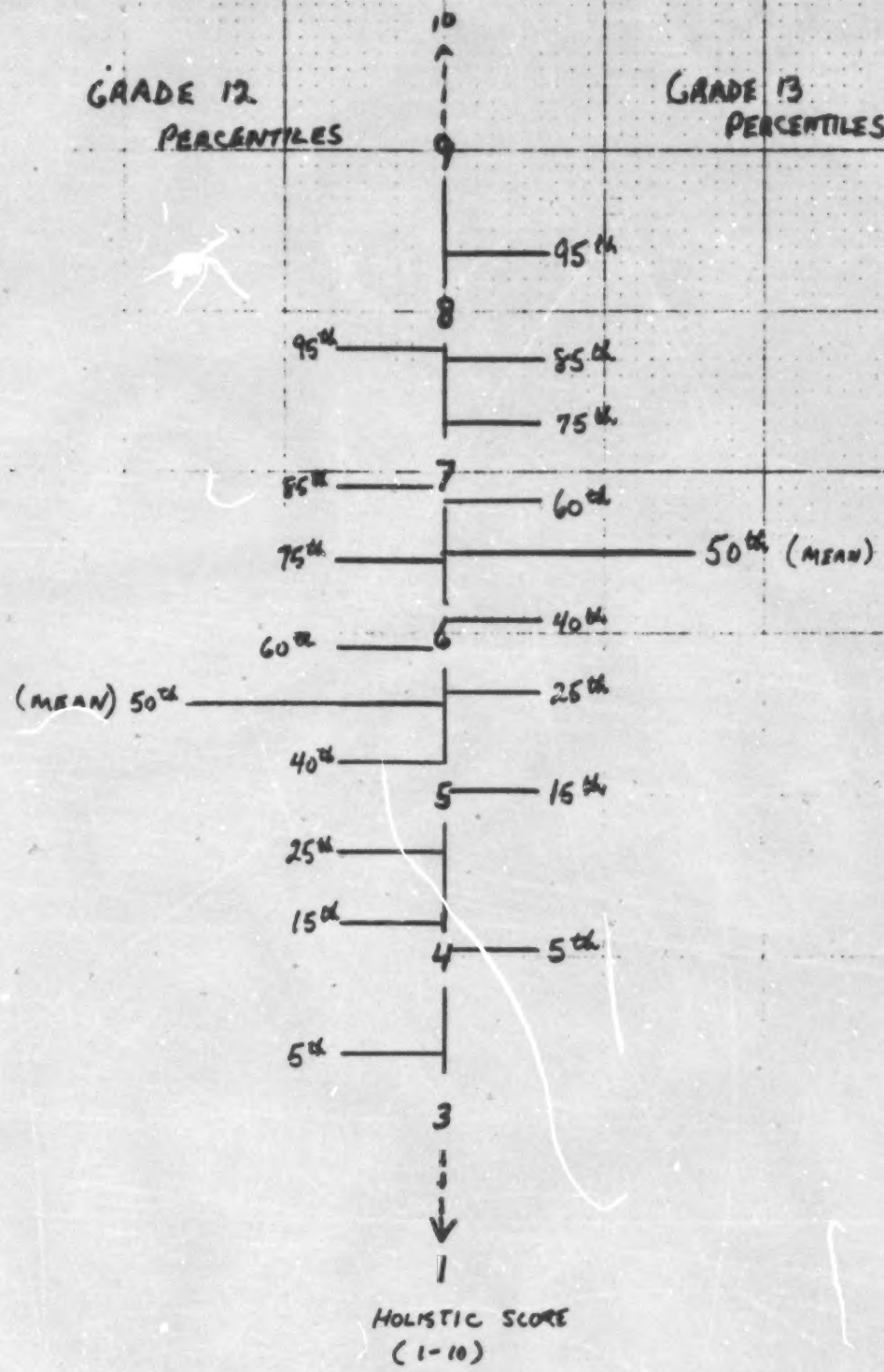
In chapter 2, I have described briefly the context of the study of student writing in the Interface Study and provided the list of topics. As noted there, the wide variety of topic choice has limited the reliability of the study as a reflection of student performance at this level. The lack of motivation for the students was another limitation, and it should be noted, in any comparison with grade 8 writing, that most grades 12-13 students did not have time to revise and recopy their compositions.

In the original Interface Report we have described the analysis (criteria and error counts) conducted on 50 of the compositions scored by all thirty-six scorers employed for the study. A number of sample compositions were also included (Traub, V.2); these are reproduced as an appendix. The Interface reports (Traub, V.2, and Russell) provide many more details concerning features of the writing; these should be consulted by teachers especially interested in the writing of students at the grades 12-13 level.

Comparison of grade 12 and grade 13 writing performance (score range 1.0-10.0) is provided in Table 4.1. The contrast between grade 12 and grade 13 performance is evident. The Interface Study which divided students into groups based on their plans beyond grades 12 or 13 showed as well that the performance level at grade 13 exceeded that of grade 12 students (Advanced Level) planning to continue into grade 13.

In the present study we have drawn a much larger sample and reported performance chiefly with reference to grade 12 score distribution, with grade 13 Quartile 1, mid-range (40-60 percentile band), and Quartile 4 performance presented for limited comparisons. For this larger sample the analysis as described in this chapter has been freshly undertaken.

Table 4.1/Comparative Distribution of Grade 12 and Grade 13 Scores: Writing



Analysis of the Sample: (A) Procedure

To the original 50 sample essays for which we had highly reliable scores, we added another 150 (2 of these later deleted because they photocopied poorly). These additional essays were drawn at random so we have confidence that they are reasonably representative of the original set of about 1600 essays. We cannot have the same confidence in the scores of the 150 as in the original 50 as reliability of scores averaged from three scorers is estimated at 0.74.

These 198 essays constitute the sample for our analysis.

The essays were scrambled so that scorers would not be treating them in rank order, and the second scorer's pile was put in reverse order to the first scorer's. All score identification was removed.

A team of experienced teachers was employed locally to perform the three analyses: analytic scoring on positive characteristics, error counts (this time on the total essay), and syntactic maturity. On the first and second analyses, two scorers were used and scores refereed or averaged. For syntactic maturity a single scorer was used.

The procedures followed (with the exception that different positive criteria were employed in the analytic scoring) were almost exactly those described in considerable detail for the grades 7-8 analysis. We used for error counts the classifications and instructions employed for grades 7-8 exposition.

Analysis: (B) Analytic Scoring for Positive Characteristics

In the original Interface Study, the criteria used for analysis were modifications of those identified by Paul Diederich* from an earlier study as those most significant to trained evaluators of writing: organization, ideas (modified here to "effectiveness of argument" to fit the mode more exactly), sentence style, diction, and flavour. These modified criteria will be found in Russell, pp. 125-127. They remain very general owing to the wide variety of topics.

For our present study, we reduced the five to four, somewhat more precise, for use on the five-point scale. The criteria appeared as follows. Note the choice for diction.

---

*_Measuring Growth in English_, N.C.T.E., 1974.

1. *ORGANIZATION*

High
- *Evidence of a plan that is clear almost from the outset.*
- *Ideas are developed in a logical sequence, beginning with an effective introduction leading to a reasonable and clearly stated conclusion.*
- *Material (pros vs. cons, or less important vs. more important) receives a quantity of treatment appropriate to its importance.*

Medium
- *Basic outline clear but somewhat faulty sequencing either because ideas are not in their most effective relationship to each other or because key steps or ideas are missing.*
- *Some imbalance because too much is given to less important ideas.*
- *The introduction may be no more than a rephrasing of the original assignment.*
- *A conclusion that, if not altogether satisfactory, makes a clear ending.*

Low
- *Outline unclear. Phases of the topic are not adequately separated; contradictory or unrelated ideas may be found in the same paragraph.*
- *Essay either trails off to no conclusion or ends abruptly with no summation.*

2. *ARGUMENT*

High
- *Viewpoint clearly stated.*
- *Each main point substantiated with illustration, examples, logic, analogy, contrast, or comparison.*
- *Some provision made for alternative viewpoint and reasons supplied for the holding of that position.*
- *The writer reveals some commitment to his position.*

Medium
- *Clearly stated viewpoint.*
- *Some but not all main points substantiated as above.*
- *Inadequate attention given to counter-arguments or too much attention given so that, ironically, the weight of the evidence shifts against the student.*
- *Transitions between stage of argument may not always be clear.*
- *The writer reveals little commitment or interest in what he is saying.*
- *A few gaps in the argument, perhaps.*

Low
- *The argument, such as it is, rests almost entirely on the writer's opinion with little explanation given.*
- *Few facts are given in support. Poor relationship, when they are provided, of facts or illustrations to what has been argued.*
- *Large and obvious holes in the case ignored.*
- *The writer is obviously bored and uninvolved in what he is saying.*

3. *LANGUAGE CHOICE SUITED TO WRITER'S PURPOSE AND INTENDED AUDIENCE*

High
- *Appropriate level of formality or informality is maintained consistent with subject, purpose, audience.*
- *Language choice is precise and clear.*

- *Occasional choice of words and images calculated to maintain interest on reader's part.*

Medium - *As for high, except that the writer occasionally lapses and word choice becomes either too grandiose or too vague.*

Low - *The writer is not always consistent in maintaining an appropriate level of formality or informality.*
- *Word choice is occasionally vague or ambiguous, often hackneyed.*

or

WORDING

High - *The writer uses a sprinkling of uncommon words or of familiar words in an uncommon setting. He shows an interest in words and in putting them together in slightly unusual ways. Some of his experiments with words may not quite come off, but this is such a promising trait in a young writer that a few mistakes may be forgiven. For the most part, he uses words correctly, but he also uses them with imagination.*

Medium - *The writer is addicted to tired old phrases and hackneyed expressions. If you left a blank in one of his sentences, almost anyone could guess what word he would use at that point. He does not stop to think how to say something; he just says it in the same way as everyone else. A writer may also get a medium rating on this quality if he overdoes his experiments with uncommon words: if he always uses a big word when a little word would serve his purpose better.*

Low - *The writer uses words so carelessly and inexactly that he gets far too many wrong. These are not intentional experiments with words in which failure may be forgiven; they represent groping for words and using them without regard to their fitness. A paper written in a childish vocabulary may also get a low rating on this quality, even if no word is clearly wrong.*

4. *EFFECTIVE SENTENCE STRUCTURE*

High - *Evidence of variety (order, structure, length) related to purpose or effect.*
- *Some ear for sentence rhythm; occasional use of parallelism, antithesis.*
- *Inessential ideas subordinated and compound sentences are infrequent unless with a purpose.*

Medium - *The student varies his sentences, though the variety may sometimes appear accidental, or simply for the sake of variety. He may tend to abuse various exclamatory devices (exclamation mark, dashes, rhetorical question) merely to "perk up" the writing.*
- *The student has not yet mastered effective subordination (through word, phrase, clause) so some sentences tend to ramble, or occasionally unrelated elements intrude.*

Low - *Too many choppy sentences and/or rambling compound or compound-complex sentences tacked together with conjunctions.*

Any variety appears haphazard.
- *The student has not developed the ability to subordinate less important ideas; consequently, the reader has difficulty determining what is important.*

We provided an initial training session and in it we requested scorers to try to use all five points on the scale. In all other respects, the procedure was as described for the grades 7-8 writing in Part 3 of chapter 3.

Table 4.2/Grades 12-13 Positive Characteristics

Grade 12 Percentiles

%ile Band	Criteria: Organization	Argument	Language Choice	Sentence Structure
Quartile 1	*3.70*	*3.53*	*3.45*	*3.44*
95-100	4.23	4.13	4.05	3.98
85-95	3.50	3.25	3.28	3.25
75-85	3.50	3.40	3.17	3.23
60-75	3.14	3.13	2.97	2.72
40-60	*3.24*	*3.06*	*2.82*	*2.94*
25-40	2.22	2.14	2.25	2.08
15-25	2.31	2.31	2.14	2.06
5-15	2.75	2.38	2.50	2.46
0.5	2.50	2.25	1.44	1.88
Quartile 4	*2.49*	*2.32*	*2.11*	*2.14*
GRADE 13	N	*(GRADE 13 DISTRIBUTION)*		
QUARTILE 1	32	4.00	3.81	3.81	3.75
40-60 BAND	36	3.40	3.35	3.15	3.11
QUARTILE 4	73	2.59	2.43	2.32	2.34

We should note the same reservations here as have already been made for the five-point scale for the grades 7-8 compositions. As scorers avoid "5" and "1" and as there is no 0-1 interval and we must take averaging into account as well, scores over 4.0 will be rare for any group of essays, and a "2" must be considered the level of unacceptability.

The patterns for all criteria present striking similarities. On all, there is a quick dip in quality between compositions in the top (95-100th percentile) band and the band just below. Secondly, we find an upturn in quality quite sharp in three of the four cases, about the 20th percentile before the final dip or collapse in the lowest band. Comparison with the error count data (Tables 4.3 and 4.4) reveals a sharp increase in error frequency about the same point: a reasonable explanation may be that most essays in the bottom quartile are reasonably satisfactory on positive features but "fail" because of too many errors. On the other hand, essays in the 25-40 band are poor from the standpoint of the criteria under examination, but "survive" in the score distribution because error frequency is tolerable.

Organization stands up better across the range and particularly in the top half of the range than do language choice and sentence style, a confirmation of my "dreariness" criticism of the essays as a whole. For the most part, we can say, however, that quality of organization of grade 12 writing is good, or at least adequate.

On this analysis, the evidence is that students need to focus more on writing style: to write with commitment and vigour. It is unfortunate that the writing stimulus situation was so poor. Likely our picture of the students' capacity to respond to a challenge is seriously biassed thereby.

Analysis: (C) Error Counts

In the present study, we have counted all errors reporting frequency per hundred words. The instructions for the counts were virtually the same as the instructions for the grade 8 essays. Again two scorers were used and counts averaged. The explanation of the code in the following tables (4.3 and 4.4) appears in chapter 3, page 32.

Just as we observed generally for the grades 7-8 writing, errors in conventions exceed all errors in sentence structure, grammar, and diction. Only in the 0-5 percentile band do we find an exception to this pattern,

Table 4.3/Grades 12 and 13 Writing: Error Counts per hundred words

(Grade 12 Percentiles)

Band	N	Aver. Words	Conventions					Sub-Total	Sentence		Grammar				Sub-Total	Total
			Sp	Ap	Pc-E	Pc	Cap		Maj	Min	Vb	Pro	Oth	Diction		
95-100	20	497	.46	.08	.08	.79	.06	1.47	.17	.12	.07	.10	.12	.22	.80	2.27
85-95	30	469	.60	.09	.05	1.02	.08	1.84	.26	.17	.18	.12	.14	.34	1.21	3.05
75-85	24	473	.70	.16	.08	.78	.08	1.80	.20	.11	.11	.17	.13	.40	1.12	2.92
60-75	32	478	.75	.19	.12	1.02	.13	2.21	.41	.16	.20	.20	.20	.38	1.55	3.76
40-60	36	415	.71	.21	.14	1.21	.09	2.36	.41	.16	.16	.24	.22	.50	1.69	4.05
25-40	18	419	.78	.20	.11	1.23	.17	2.49	.52	.25	.23	.35	.31	.52	2.18	4.67
15-25	18	341	1.65	.26	.22	1.17	.21	3.51	.72	.29	.29	.37	.35	.55	2.57	6.08
5-15	12	325	1.87	.28	.12	1.04	.18	3.49	.79	.26	.54	.31	.29	.64	2.83	6.32
0-5	8	435	1.73	.37	.10	.95	.12	3.27	1.61	.24	.63	.36	.39	.96	4.19	7.46

Table 4.4/ Error Counts

Summary: Grade 12 Bands

Band	N	Aver. Words	Conventions Sp	Ap	Pc-E	Pc	Cap	Sub-Total	Sentence Maj	Min	Grammar Vb	Pro	Oth	Diction	Sub-Total	Total
Quartile 1	74	477	.59	.11	.07	.87	.07	1.71	.21	.14	.13	.13	.13	.33	1.07	2.78
			25%		34%			62%	13%			14%			38%	
40-60 %ile	36	415	.71	.21	.14	1.21	.09	2.36	.41	.16	.16	.24	.22	.50	1.69	4.05
			23%		33%			58%	14%			15%			42%	
Quartile 4	38	356	1.73	.30	.16	1.08	.18	3.45	.97	.27	.45	.35	.34	.68	3.06	6.51
			31%		19%			53%	19%			18%			47%	

Summary: Grade 13 Bands

Quartile 1	32	480	.42	.06	.07	.74	.06	1.35	.19	.15	.09	.09	.11	.24	.87	2.22
			22%		36%			61%	15%			13%			39%	
40-60 %ile	36	473	.66	.13	.09	.87	.10	1.85	.26	.11	.15	.18	.13	.38	1.21	3.06
			26%		31%			60%	12%			15%			40%	
Quartile 4	73	382	1.22	.25	.14	1.10	.15	2.86	.71	.24	.32	.34	.30	.60	2.51	5.37
			27%		23%			53%	18%			18%			47%	

a dramatic one in which error frequency in conventions slightly declines and major errors in sentence structure double from the count in the previous band (1.61 per hundred words compared to 0.79).

It is evident again that in general the scores decline as error frequency increases, though this pattern is not characterized by the averages in the upper bands of the grade 12 distribution. In the Interface Report correlations between score and frequency of several classes of error were carried out on the original sample of fifty. (Table A1.8, p. 117, Traub V.2) These correlations, all negative, were as follows: Spelling, .52; Total Conventions, .43; Total Sentence Structure, Grammar, Diction, .50; and Total Errors, .58. While it is evident that error frequency, as might be predicted, has an influence on the score awarded an essay, the influence is not as great as one might expect. Other characteristics of writing quality influence scores as well.

For three-quarters of the distribution (25-100th percentiles, grade 12), the average increase in error frequency as scores decline is gradual; about 60% of the errors are in conventions. Errors in capitalization are so infrequent that, as we saw for grades 7-8, the count is scarcely worthwhile.

Across the range above the 25th percentile grade 12 (i.e. above the 10th or 15th percentile grade 13), the frequency of major sentence errors is encouragingly low. For the mid-range grade 12 student, it approximates two errors in 500 words, and for the mid-range grade 13 student about one for 400 words. When we recall the old grade 13 rule of thumb "three major errors and you fail," we may say that a high proportion of grade 13 students today would certainly not fail on this measure.

As well, grammar errors are relatively few. All errors in verbs - subject-verb agreement, tense sequence, and principal parts (inclusive of misspelling of principal parts) - come to about 1½ per 1000 words in the mid-range essays for either grade 12 or 13.

Pronoun errors are also infrequent, though somewhat higher than we find in grades 7-8. Does this mean that the student has un-learned rules for pronoun agreement? Probably not. Rather, the more complex syntax in senior writing (e.g. more distance between the pronoun and its reference, more frequent use of the indefinite pronoun) renders it more liable to errors of certain types. Raw figures appearing to reflect trends in error frequency can mislead if the evaluation leaves out of account the

types of structures students are using at this level.

In any event, in all parts of the range, grammar errors account for a low percentage of the total errors in the compositions.

The foregoing observations concerning grammar and sentence structure need to be stressed in the present climate of distrust of school achievement. It is just not the case that secondary school graduates cannot write grammatically or cannot write a sentence. Criticisms of that sort are likely to influence program development in directions which contradict the data. In saying that the student can write a sentence, we are not saying the student writes with stylistic excellence. Perhaps, as already noted for grade 8 writing, there should be more work in development of stylistic competence rather than on correctness drills.

We have already, in chapter 3, issued abundant warning about the punctuation count, the single most frequent source of error for most of the bands. First, we counted with some ruthlessness, including as errors many punctuation "faults" some good writers today would consider a matter of stylistic choice. Second, punctuation conventions do change rapidly with the trend currently towards reduction of internal punctuation. Third, we have not included under conventions punctuation errors producing actual sentence errors; these are all classified under "Sentence." A reasonable route to improvement of minor punctuation again seems to tie back to study and practice in effective sentence style rather than imply more punctuation drills on the "eighteen rules for the comma."

Also we have noted in chapter 3 that any summary of data obscures information. Consider, for example, spelling in the 40-60 band for grade 12. There were 36 sample essays in this band with an average length of 415 words. When frequency is averaged for these 36 essays we obtain a figure of 0.71 per hundred words. However, essay-by-essay examination reveals that seventeen of the errors occurred in one essay. Two other essays had more than five. Eighteen had two or fewer errors and, of these, three had none.

We observe a sudden jump in frequency of errors in verbs between the 15-25 band and the 5-15 band (from 0.29 per hundred words to 0.54). When, however, we examine the essays in the 5-15 band one by one, twelve of a total of twenty-one errors found occurred in a single essay. Remove that essay and we find a total of nine errors in about 4000 words.

Table 4.5/Syntactic Maturity Grades 12-13

Grade 12 Distribution

%ile Band	N	Sentence: Average Length	T-Units: Average Length	% Compound	Noun Cl. per Essay	Sentence Introduced by: Phrase As	Clause %	Total	Aver. No. Para	Aver. Length Para. (words)
Quartile 1	*74*	*19.3*	*16.2*	*11.0*	*3.3*	*15.8*	*7.9*	*23.7*	*6.2*	*82*
95-100	20	19.6	16.2	11.4	2.5	17.8	7.2	25.0	6.1	90
85-95	30	19.6	16.5	10.6	3.7	16.8	7.5	24.3	6.3	82
75-85	24	18.5	15.8	11.1	3.4	12.9	8.8	21.7	6.4	76
60-75	32	20.6	16.2	10.8	3.4	12.7	7.3	19.9	6.3	77
40-60	*36*	*19.7*	*15.8*	*10.0*	*3.6*	*10.6*	*11.7*	*22.2*	*5.3*	*78*
25-40	18	20.6	16.5	11.5	4.9	8.2	11.7	19.8	5.8	91
15-25	18	17.9	14.7	10.1	3.2	9.2	7.9	17.1	5.7	75
5-15	12	17.8	14.8	10.8	1.9	9.8	5.7	15.4	5.1	76
0-5	8	24.1	14.9	9.6	1.6	11.3	5.8	17.0	4.0*	82
Quartile 4	*38*	*19.2*	*14.8*	*10.2*	*2.5*	*9.8*	*6.8*	*16.6*	*5.2**	*77*

* The last ranked essay contained 22 "paragraphs." It is omitted from the number and length of paragraphs count as are two essays that went unparagraphed.

We must be very cautious, therefore, in looking at any band and saying "Students in such and such a range in grade 12 or 13 make X or Y errors in sentence structure, spelling, or anything else." Our tabular information is useful in providing us with general data - types of error and relative frequency or infrequency. It is wrong to conclude from it that "most students" and certainly any particular student will or does perform in a particular manner in his writing.

Analysis: (D) Syntactic Maturity

The analysis of syntactic maturity was conducted in exactly the manner described in the previous chapter for the grade 8 writing.

We again counted paragraphs and averaged paragraph length, but, as noted in chapter 3, we doubt that the information is particularly useful. Two essays of the 198 in our sample went unparagraphed, and one (the essay with the lowest holistic score) had twenty-two "paragraphs." Evidently almost all senior students do use paragraphs.

The summary of information for grade 13 appears here as Table 4.6.

Table 4.6/Grade 13 Summary
(Grade 13 Distribution)

Band	N	Sentence Average Length	T-Units Average Length	% Compound	Sentence Introduced by Phrase As %	Clause As %	Total
Quartile 1	32	19.4	16.0	11.5	16.0	7.3	23.4
40-60 %ile	36	19.8	16.4	9.8	13.3	9.1	22.4
Quartile 4	73	19.5	15.5	10.0	9.2	8.9	18.1

As for the grades 7-8 writing, we find little difference in mean sentence length in relation to score except (and here the number of essays in the sample is small) we do find a sudden jump in the worst essays (0-5 percentile band) comparable to the jump in the grades 7-8 Narrative in the same band, evidence perhaps that one characteristic of the weakest writers may be sentence structure completely out of control. We must stress "may" here, as the phenomenon did not occur in the grades 7-8 Exposition.

A difference of 1.4 is found in average T-Unit length between essays in Quartile 1 and Quartile 4 in the grade 12 distribution, suggestive that this index of syntactic maturity is not useful in comparing general quality of writing at the same age/grade level. Its best use in "summative" evaluation appears to be comparison of performance at different grade levels. In this instance, we observe (Table 3.5) mean T-Unit length in the 40-60 band (grade 8) to be 12.5 for Narrative and 13.1 for Exposition: process/procedure. For grade 12 mid-range Exposition: point of view, the figure is 15.8 (Table 4.5) and, for grade 13, 16.4 (Table 4.6).

A more detailed discussion of the T-Unit as a measure of syntactic maturity appears in chapter 6 of the original report to the Ministry. In American studies cited in Stewart's report,* words per T-Unit are given as grade 7 — 9.8, grade 8 — 11.3 and 9.8, grade 12 — 11.3. Stewart's New Brunswick study produced means of about 13 for grade 12 "upper level" students and of about 11 for the "regular" (pp. 40-41). Each study had its own context so that it is difficult to compare usefully. On the surface it appears that the students in our study compare quite favourably.

The compound sentence count revealed almost no difference across the score range.

We hoped that the frequency with which a sentence was introduced by a phrase or a clause might be a useful measure for distinguishing between more and less effective writing in terms of effective sentence style. The quartile averages in Tables 4.5 and 4.6 support our hope, and, as it is a count relatively easy to make, it may be useful to teachers in the classroom as well as to systems interested in assessing general performance in writing. The difference between mean scores in the upper quartile and the mid-range is very small for either grade 12 or grade 13 writing, but the fourth quartile scores are considerably lower. A further distinction is evident when we compare relative frequency of introductory clause over introductory phrase in different parts of the range. In the three bands reported for Quartile 1, the clause appears about twice as often as the phrase. Over the mid-range, we find a trend towards equal frequency, with a return to an approximation of the 2:1 ratio in the worst essays — an interesting "index" perhaps worth further study.

*Previously cited, *Research in the Teaching of English*, vol. 12, no. 1, p. 38.

Afterword

With a few fascinating exceptions, especially when the student decided not to take the test seriously and to have a bit of fun, the essays at grades 12-13 lacked imagination and vitality. The extent to which this was the result of the conditions under which the sample was taken and the extent to which this is a normal characteristic of students at this level we cannot, unfortunately, determine from our present sample.

As well, as noted in chapter 2, lack of control of the topic variable makes useful generalizations rather difficult. The following observations may, however, be made:

- Generally the quality of organization is satisfactory to excellent for most students in grades 12 and 13.
- The quality of sentence structure is less satisfactory in general, though serious errors in structure are reasonably infrequent above the 25th percentile grade 12 and 10-15th percentile grade 13.
- The frequency of use of the introductory phrase or clause as an index of quality of sentence structure or style appears to be useful. The better quality of essay, judged by the holistic score, employs such sentences with greater frequency than do the weaker essays.
- Convention errors are generally more frequent than all others, and grammar errors in particular are far less frequent than would be expected from the hammering English writing competence has been given by the public.

As part of the Interface Study, a number of exemplar essays were provided in a part of the report (Traub V.2) which was not made generally available. I have reproduced these essays as an apprendix to this monograph. The sample is small but may assist in providing some impression of the writing quality. The comments are mine as they appeared in the original Interface report.

5 Development in Writing Ability: Grades 8-12

Numerous practical limitations within each of the Intermediate and Interface Studies of writing impose severe constraints upon any effort to make a formal comparison of writing performance of grade 8 and grade 12 students in Ontario. Within this study, we must be content with a few general observations and examination of those parts of the data which are at least suggestive of trends and which appear helpful in pointing the way towards future studies in which appropriate comparisons might be formally undertaken.

Our chief limitations here, all of them remediable in any future study on a school system level or on a larger scale, are these:

- Lack of control of the topic variable in the grades 12-13 sample
- Lack of comparability of mode: The grades 12-13 writing was not in the same mode as either of the grade 8 assignments
- Problem of incentive for students writing in the Interface Study
- Non-representativeness of the sample: The grade 8 students were all from schools and systems volunteering to join the project

In spite of these limitations a number of general observations appear reasonable.

1. Between both grades 7 and 8 and grade 12 Advanced level and grade 13, scores indicate considerable improvement.

Both studies suggest strongly, therefore, that there is improvement in the quality of student writing as the student proceeds through the system. The degree to which this is the result of maturation and the degree to which school programs have influence is, of course, a separate question.

2. In both groups, grade 8 and grade 12, errors in conventions (the courtesies of language) exceed all classes of grammar, syntax, and diction errors combined, and across most of the score range (i.e. for the large majority of students at either level), frequency of grammar errors is low relative to all other errors. For the mid-range at grade 8, we find

a frequency of 0.74 per hundred words for Narrative and 0.34 per hundred words for Exposition — one grammar error in 300 running words. At grade 12 mid-range, the frequency is 0.62.

As we have seen in comparing counts for grade 8 Exposition and Narrative, the mode somehow has strongly influenced error frequency counts; and, as we have only one "mode" but great topic variety at grade 12, we cannot make a generalization about any "absolute frequency." On the surface, it appears that grammar errors are about as frequent at both levels, and here an important additional caution about comparing data must be introduced and underlined.

3. Comparison of error frequency data at the two levels may seriously mislead as the influence of mode reveals. Additionally, any judgments about differences or their apparent lack must be made in the context of the complexity or "syntactic maturity" of the writing.

On the simplest level, assume examination of writing at two different age levels revealed about the same error frequency in spelling. The conclusion might be "Those kids don't improve in their spelling after four additional years of schooling," a conclusion which, though understandable, is likely wrong. If we assume an increase in active vocabulary between grade 8 and grade 12, then the occasions for spelling errors may be very different. Many students in grade 12 may have mastered many of the spelling problems found in grade 8 writing and be having problems instead with more difficult, more esoteric vocabulary.

On a more complex level, if we find that grammar and sentence structure errors do not, on a simple "count," decline very much between grades 8 and 12, we must, before condemning the school system, determine whether students at grade 12 use more complicated structures — more intervening words between subject and verb, more collective nouns, more clauses, more situations; in other words, where errors of one sort or another are more likely to be frequent. Though both errors would be classified the same way in a count, there is a difference surely between "One of the girls were..." and "One of the girls who were seen at the party with many of their friends were..." In the latter, it is a much longer trip from the bare subject to the verb.

We do find (again 40-60 band grade 8 and grade 12) no striking differences between grade 8 and grade 12 about which to break into wild cheers:

40-60 BAND

	8 Narr.	8 Exp.	12
Major Sentence	0.96	0.42	0.41
Grammar	0.74	0.32	0.62
Spelling	0.84	1.07	0.71
Minor Punctuation	1.25	2.45	1.21

We do know from the syntactic maturity counts undertaken that there are differences in sentence complexity which likely account for the appearance of little difference in error counts. For example, we observe a difference in average T-unit length in the grade 12 over the grade 8 writer in these same bands:

	8 Narr.	8 Exp.	12
Average T-unit (40-60 band)	12.5	13.1	15.8

And we also find an increase in frequency of use of the noun clause:

40-60 BAND

	Grade 8			Grade 12		
	Aver. Essay Length	No. per Essay	No. per 100 Words	Aver. Essay Length	No. per Essay	No. per 100 Words
Narrative	328	0.7	0.21			
Exposition	187	0.7	0.37	415	3.6	0.87

From these differences in performance in measures of syntactic maturity, both suggestive of growth between grade 8 and grade 12, it is reasonable to conclude that the absence of large declines in error frequency, rather than suggesting that performance on "grammar" and "sentence structure" is unchanged, suggests otherwise - that there is growth and improvement in the handling of English grammar and syntax.

Though our grades 12 and 13 populations differ, it is well to remember that we do observe a substantial improvment in mean performance as well between grades 12 and 13.

4. To the extent that we can consider the average T-unit length a significant measure of increased writing maturity or sophistication, we have strong evidence from our syntactic maturity tables in chapters 3 and 4 that student writing does mature significantly between these levels.

5. Despite the problem created by the very different modes of writing at the two levels, a general observation concerning "organization" may be made.

What we may classify on our analytic scores as a "satisfactory" performance for organization characterizes much of the grade 12 writing above the 40th percentile, and we find reasonably satisfactory averages in some of the lower bands as well. The grade 13 distribution reveals an even more impressive performance on this measure. In grade 8, however, the quality of organization drops on our scale below 3.0 in the 40-60 band (Narrative) and 60-75 band (Exposition), suggestive that senior students generally are considerably more adept in organizing effectively whatever they wish to say.

6. Both error frequency tables and positive characterization of the writing point out to sentence <u>style</u> and <u>structure</u> frequently ineffective in achieving an intended effect or purpose. This finding implies the need for greater attention in language and writing programs in providing good sentence models.

6 A Program for Writing Evaluation within a School System

The earlier chapters have described the Intermediate Evaluation Project in some detail, as well as the sampling and study of writing performance near the conclusion of elementary grades and near the conclusion of secondary school. The data concerning performance, even in these limited studies, will, I hope, be of considerable help to teachers and curriculum planners in obtaining a more exact picture of how students at these stages write as a basis for general program planning and emphasis and as an information base in responding to criticism from the public at large.

In this chapter, I shall outline a workable approach to writing evaluation, a plan for systematic evaluation that might reasonably be undertaken by any school or school system, either across one grade level or across several grade levels from grades 6-13. The model is derived chiefly from the Intermediate Evaluation Project model which incorporated program planning, a good sampling of writing over the term reflective of program emphasis, and means for improved dialogue between teacher and student, together with the summative evaluation through the final pair of assignments holistically scored and subsequent analyses of a sample representing the full score range.

Analysis of the writing in the Interface Study was entirely a summative evaluation exercise having purposes related to the much larger framework of performance, secondary-postsecondary program match, and the relative value of a variety of achievement measures.

And even an exclusively "summative" evaluation has a good deal of value to systems wishing general information for appropriate grade levels concerning performance in writing and trends in performance from year to year. Such an evaluation plan does not require the testing of every student, but of a random sample. That is its advantage. Its disadvantages, conducted separately from program development and not involving

every student, are several. There is no feedback of information to the student, (consequently) little likelihood that the student will be committed to the exercise and put forward his best effort, and the danger that impact upon program development with teacher involvement and commitment will be slight - not unlike the present situation in many systems where a battery of tests is administered annually and automatically with little practical use made of results.

Therefore, and because, according to teacher response (see chapter 2, pp. 6-7), our Intermediate Project did produce many desirable effects, I strongly recommend a plan for writing evaluation that engages teachers and students from the beginning, providing information useful for individual student diagnosis, information concerning development of writing ability through the year (formative evaluation), and summative information on performance. And this summative evaluation provides reasonably reliable information to the student concerning where he stands in the group, to teachers and curriculum consultants for program reassessment, and to the public concerning the status and trends in writing performance. The plan reasonably meets the needs of all clients in the enterprise.

Teachers in the first year of the project were participants in every sense. It was they who identified the modes of writing deemed most important for grades 7-8, who selected many of the topics which became the package of term assignments, who identified the criteria employed to evaluate the writing over the term, and finally it was they (with the work distributed among all) who refined those criteria as they now appear in the Criteria Handbook. That strong participation from the outset likely accounts for the wide, favourable response concerning the effect of the project on writing programs - and upon student responsiveness to the writing program.

STAGE ONE

The first stage, surely, in any plan for evaluation of a program, is the achievement of reasonable consensus concerning its main objectives so that evaluation strategies designed or chosen will be in conformity with the program. Some readers may observe, "That goes without saying," but in fact all too few evaluation efforts include this step.

In the evaluation of writing, the necessary first stage is the identification of the modes or types of writing (with some consensus concerning

reasonable length) considered most appropriate for the particular grades involved.

In the Intermediate Evaluation Project, because so many teachers and systems geographically diverse were involved, this had to be established by survey employing rather general terms (first-person narrative, exposition: process/procedure, etc.). The package of term assignments subsequently developed (each teacher to choose up to four of the seven modes, as long as at least one was expository) was based upon the seven found to be most generally popular among the grades 7 and 8 teachers, with the final pair of assignments provided in modes for which there was almost universal agreement.

Individual school systems undertaking development of a writing evaluation plan should be able to define the modes of writing (and their characteristics or criteria) more precisely than was possible in this wider-ranging project. Some systems introduce and emphasize "task" writing earlier than others, for example. In reviewing modes of writing deemed appropriate (e.g. in grade 7 in contrast to grades 10 or 12), I strongly recommend study of James Moffett's *Teaching the Universe of Discourse* and James Britton et al., *The Development of Writing Ability 11-18*. The program should reflect increasing sophistication in mode and style over the grades with the more abstract forms of writing not introduced too early. Our observation that process/procedure writing was generally inferior to narrative at grade 8 carries with it a message, but the message is not necessarily that this mode should therefore receive increased emphasis at that relatively early stage of the writer's development.

Differentiation of characteristic modes is a gradual development. In early grades writing tends not to be definable by mode, and it is chiefly for this reason that I do not recommend a formal writing evaluation plan of the present sort below grade 6. It would be inappropriate in early stages of writing development to impose a conformity to various criteria.

STAGE TWO

Stage Two follows hard on the heels of Stage One. As modes are identified and certain expectations concerning writing performance emerge, teachers should begin to identify and then to specify criteria. Some of these will be general criteria, applicable to most writing assignments; others should be mode-specific.

The task of elaborating criteria is a large one. At the senior level, teachers will find the criteria elaborated in Paul Diederich's *Measuring Growth in English* very helpful as a beginning. Those with some modifications we used in the Interface Study (see pp. 51-52).

The Criteria Handbook, developed with the cooperation of the teachers involved in 1976-77, will be a helpful resource for teachers across the Intermediate Division. Though the criteria were originally elaborated for grades 7-8 writing, the "high" version of most (somewhat idealistic for grade 8 anyway) with few modifications will be found useful for grades 9 and 10 as well. I commend this collection for two reasons: first, it is the work of a lot of practising teachers in Ontario, and, second, it will enormously reduce the chore for teachers in elaborating their own. The one "criterion" I would demand of all criteria is that they be rephrasable for ease of use with students. If the students don't understand them, the game is lost.

Samples of criteria from the handbook appear, with small variations, in chapter 3 as the criteria used in the analysis of our sample essays. Others appear (one "general" and one "specific") on pp. 112-13 of *Intermediate Division English*, 1977 (Ontario).

STAGE THREE (Use of Criteria in Both Diagnostic and Formative Evaluation)

While by no means all writing, especially experimental creative efforts, which the student produces over the school year, should be subjected to formal evaluative scrutiny, reasonable accord should be established concerning those modes to be evaluated, the frequency of that evaluation, and at least some of the criteria to be used. Student contribution to the establishment and clarification of criteria should be welcomed! If the kids can articulate what they believe is important in their writing, you are far ahead.

Within the framework of the Project, participating teachers were required to mark four assignments over the term employing analytic scoring and to submit a combined mark for these four as a term mark. In the computation of the final score for the student, the term mark accounted for 60% of the total, the holistic score on the final pair of assignments for 40%. An assignment package of seven modes with alternate sets of topics for all but one (i.e. fifteen subpackages from which the teacher could select) was prepared and distributed to all participating teachers with three

basic recommendations: to distribute these four formal assignments over the year, to employ analytic scoring, and to select or even re-select modes each deemed especially important in his/her program. Teachers were also given early information concerning the two modes that would be assigned universally in April.

This schedule and degree of flexibility was for the Project the best fit between the need for a consensus where a large number of systems and students were involved and the desirable idiosyncrasies of individual school programs. The framework constructed by an individual school or system, especially if several grade levels were to be involved, would have differences: some modes would be emphasized at particular grades and perhaps not appear in others, the frequency of these "formal" assignments might be increased, modes and assignments might be more sharply specified, and particular criteria underlined.

A plan combining diagnostic, formative, and summative evaluation requires a consistent approach to the maintenance of individual files for student writing and design of simple checklists, tailored somewhat to individual performance and problem areas, so that diagnosis can be made and evidence provided that improvement is taking place. Hence, in the *Intermediate English Guideline*, pp. 110-11, the recommendation of the writing folder and the use of the writing checklist.

Whether the teacher (or, in group marking efforts, students) are "scoring" for more mechanical concerns or for "organization," "diction," and so on, defined for a particular mode, no more than <u>five</u> features should be considered in analytic scoring at any one time. The Guideline illustrates use of a 5-point scale with "1" as weak and "5" as strong. Performance expectations should be scaled to fit reasonable achievement levels of the class - room for growth, but not scaled so high that the students will become discouraged. There is nothing magical about a 5-point scale, but it does provide flexibility and a reasonable range of performance. Teachers may wish to add a total impression mark to the analytic scores.

Use of an analytic scale with criteria clearly defined for both teacher and student will increase <u>somewhat</u> the consistency of scoring and treatment of writing. However the main purposes of this scoring approach are clearer identification of the appropriate features of writing particular modes called for with greatly improved feedback to the student concerning

his performance and specification of writing criteria reasonably to be met at different stages in the year and at particular grade levels.

I do recommend to the teacher, whenever feasible, to use this scoring approach on other assignments besides the formal ones as a means of providing consistent feedback, and, where several teachers of a grade level have agreed on the formal assignments and criteria, to exchange compositions for scoring. Practice scoring sessions, once criteria have been established, are valuable in themselves for clearer identification of program objectives and greater consistency from class to class in treatment of student writing.

More detailed treatment of analytic scoring and use of the approach for diagnostic as well as formative evaluation appear in *Intermediate English* and its forthcoming evaluation support document as well as in Paul Diederich's *Measuring Growth in English*.

Consistent use of analytic scoring in association with charting of performance to criteria provides, besides information concerning individual students' needs and achievements, ongoing information concerning general achievement with respect to major writing program objectives, and the teacher is in a position throughout the year to provide evidence of desired change and to revise objectives, program, or strategies as general performance suggests.

As previously noted, measures of performance in writing over the year should be combined with any measures taken at the end of the year in reporting student achievement. Writing performance must be assessed on the basis of more than one or two occasions; here, particularly, formative and summative evaluation should be combined.

I have recommended that the scale (e.g. 1-5) in analytic scoring be applied differently with different groups: where classes are streamed, a "4" or "5" on that scale should represent a higher level of performance in the group of better ability. No teacher should be required to bury whole groups of students with scores of "1" or "2" on virtually every criterion; such an approach would destroy all motivation among less able students. Of course this implies different standards for different groups.

However, most of our "clients" - students, parents, trustees, as well as teachers - have a strong and legitimate interest in a common standard

or a common meaning applied to measures of performance. And this can be well accommodated when formative measures (here expressed as a combined score on term writing) and summative measures, to be described below, are combined. The teacher's term marks are simply rescaled against the score distribution in the holistic scoring of the final assignment or pair of assignments.

This is what we did in the Intermediate Evaluation Project. The teacher submitted a term mark as a per cent based on the four formal assignments. This was scaled against the distribution of scores of his students in the holistic scoring on the whole population, and the rescaled score and holistic score combined 60:40, with the term counting 60%. The student received three scores: the rescaled term score, the holistic score, and the combined score.

STAGE FOUR (Summative Evaluation)

Summative evaluation may have one or several purposes. These include

- Provision of individual scores reliably measuring writing competence in comparison with the rest of the group (class, grade, grade level in system): as information to the student and to the staff to assist in determining promotion or program.
- Provision of information in some detail concerning the characteristics (strengths and weaknesses) in the writing of students at a particular grade level.
- Comparison of performance of students at different grade levels in order to reveal evidence of growth and characteristics of growth in writing ability.
- Before and after measures to judge the effectiveness of a present writing program and/or of an innovative writing program.

  Particularly at a time of introduction of new Guidelines in English, teachers and administrators will have an interest in knowing whether a particular approach to the teaching of writing and language is achieving the effects intended or assumed.

- Provision of annual data on performance at some stage of the program (e.g. grade 8 or grade 12), this in turn for public information and as a means of determining whether an acceptable standard of performance is being maintained.

Purposes will dictate strategies. Some of these objectives can be met by scoring and analysing the writing of a sample of students, the sample taken under controlled conditions. The "annual" check into standards would likely require burial for rescoring of a sample of writing from past administrations. Scoring reliability must be higher for measuring individual student performance. And so forth.

In the plan I am recommending, most of the purposes for summative evaluation may be achieved.

At Stage One, or at least early in the school year, participating teachers should come to an agreement about the modes of writing to appear as the assignments for holistic scoring. I recommend two assignments on different days, in different but important modes, with control of the topic variable. The first-person narrative and process/procedure assignments used in 1977 (see page 9) worked well. Though there was a good deal of freedom in choice of content, the type of treatment was clearly specified, making it possible for scorers to treat all of the writing in each mode as comparable.

Students as well as staff should have early and complete information concerning the final pair of assignments and the conditions under which they are to be written; all that is withheld is the specific topic or stimulus. Students should have adequate time to write, revise, and recopy, and have access to dictionaries, a thesaurus, and other such resources. In short, the situation should be normalized as far as possible.

The compositions should be recopied onto the type of treated paper that provides multiple copies. Photocopying has numerous problems. Student identification is to be coded in such a way that no information concerning the student, his program, or his school is betrayed to the scorers.

If three scorers are used, the original and two copies of each composition will be required. Immediately following the writing of the composition, the teacher will separate the packages into three piles, making certain that second and third copies are legible.

The next, clerical stage requires thorough advance planning: the mixing of the compositions by mode, the packaging for scorers, and establishment of the means by which scores can be conveniently recorded and reassembled and processed. A computer and key-punching facility will be important at this stage if the group of students is large or several schools are involved.

For this final pair of assignments, holistic scoring is the best available means of achieving high reliability of scoring.

Holistic Scoring

Holistic, or rapid impression scoring, has been used to good effect in numerous settings in order to rate pieces of writing in relation to other pieces of writing rapidly and with acceptably high reliability.

Of itself, it provides no analytic description of the character or quality of the writing, though this feature can be built into the process as a second-stage analysis for baseline information. Where this second stage is intended, a fairly large random sample of essays should be duplicated and "buried" for scoring by a larger number of scorers in order to obtain highly reliable scores.

For holistic scoring, it is most desirable that

- All student/school/program/grade level information be masked.
- The writing all be in one mode and, ideally, on one topic (i.e. the writing is a response to one specific stimulus).
- Scorers have a "straight run" at the essays in fairly large batches (not a few marked in a spare here and there) without distractions.
- A small number of essays be buried for scoring by all scorers (to calibrate scorers in their use of the scale).

This last feature permits one with computer assistance to bring scorers to the same expected mean - desirable but not essential.

Each composition should be rated by a minimum of two scorers; three would be better (probably the best balance of practicality against reliability), and four excellent. Increases beyond four provide very small gains in reliability. In the Interface Study, reliability for two scorers on a single piece of writing was calculated at .66, and for three .74 (with calibration, .68 and .76 respectively).

Research studies in holistic scoring indicate that a short training session for the scoring team improves reliability of the scoring. Though this step was not available to us in either the Interface Study or the Intermediate Writing Evaluation Project, it appears feasible in a school or system setting, and I strongly recommend it.

Reliability will also increase where there is a greater difference in essay quality: for example, when writing across several grade levels is being scored.

The basic instructions are very simple.

Scorers are instructed to read rapidly a number of essays rating each on

an appropriate scale using all the points on the scale. They are not required to distribute the essays evenly across the score range.

A scale of 1-10 is recommended. We suggest an even-numbered scale in order to avoid the tendency of some scorers to pile up essays at the exact midpoint. With a scale such as 1-10, the very best compositions in the batch are to receive the highest score (10); the very worst receive 1. Scorers should check occasionally to make sure that all points on the scale are being used.

In starting out, scorers should give a preliminary score to 15 or 20 compositions in order to get their range, and then rescore these once they have a sense of the total range. This will avoid the problem of having to make major adjustments later on when they find essays much better or much worse than most of those that they happen to have scored.

Markers should be informed that the scale is quite arbitrary; the points on it are not marks like percentages. A score of 4 is not a failure and a score of 8 is not first-class honours, and so on. The scale is simply a numerical device for ranking the compositions in relation to each other. A composition scored at 5 is better than, on general impression, a composition scored at 4. And that is all that the scale means.

Most teachers have initial difficulties in "keeping their hands in their pockets." They are so used to scribbling marginal notes, circling spelling errors, and so forth, that they feel somewhat frustrated. They must be encouraged not to pause for this kind of detail and not to put any marks on the paper other than the score itself. The more time they spend browsing through the paper, the less reliable their scores are going to be.

It is not easy, either, to convince teachers that holistic scoring is indeed rapid. They will need a little coaxing, but the experience will bring belief. In the Intermediate Evaluation Project last spring, teachers were rather fearful of the load involved in scoring approximately 175 compositions each. However, we and they were very pleased to find that the task was not nearly so burdensome as anticipated. In fact, many teachers reported to us that they enjoyed the experience very much, partly because it gave them an opportunity to look at student writing right across the province. This gave them a better perception of the quality of work their own students were producing.

It is important, once teachers get into the scoring, that they score a fairly large number of essays at a stretch. Each time they take a long break away from the scoring they have difficulty in finding their range once again. This means backtracking, frustration, and lower consistency.

If scorers are to be calibrated, a small number of compositions, randomly selected, should be photocopied and buried in each scorer's package.

Assuming (Stage 5, below) a further analysis is to be made on a larger sample, more compositions should be randomly drawn and photocopied for scoring by six or seven scorers in order to provide for these highly reliable scores. Then, in the subsequent analysis, the report on characteristics on writing can be related to score range with more confidence, as outlined in our reports in chapters 3 and 4.

I should note here that teacher groups will have to make practical compromises with this plan. The clerical task of pulling the sample, mixing them into bundles, and handling scorer calibration will be too much for a small group to undertake. Within a single school there won't be sufficient scorers to permit the above refinements. Nevertheless, a random sample may still be drawn and a fruitful analysis made.

STAGE FIVE (Analysis of a Sample)

The advantages of holistic scoring are its speed and accuracy; its disadvantage to the teacher and student is that it provides no information about the characteristics of the writing. For this reason, a follow-up analysis of a sample is important for provision of data useful to teachers in assessing the program.

A comprehensive analysis - positive characterization, error frequency, measures of syntactic maturity - has been described in chapters 3 and 4 for both intermediate and senior writing, and we have been able to make some judgments there about the kinds of analysis likely to be most productive of useful information. No doubt the reader has made his own judgments as well!

Teachers or administrators interested in the detailed instructions we provided to our "analysts" may consult the original report where the instructions appear in full detail as appendices. Likely the categories as defined for our tables in chapters 3 and 4 will be a sufficient guide in selecting the most useful.

The set of compositions to be analysed should before packaging have all score or ranking information obliterated and should be shuffled so that they will not appear in order of quality.

Positive Characterization: Analytic Scoring

For positive characterization, you will need two scorers for each composition, with scores averaged. In our analysis we "refereed" rather than averaged major differences (2 or more points on two or more criteria) by adding a third scorer for those essays. For this "final" analysis, I recommend a 6-point rather than the 5-point scale used over the term, as scorers tend to shy away from the highest and lowest points and the averaging effect as well tends to shrink scores into the middle of the available range.

As we have done in our analysis, I suggest selection of four criteria: organization, sentence structure, diction (all expressed to fit reasonably closely to the demands of the mode), and one other particularly related to the mode, with instructions to scorers to emphasize the positive aspects. Errors will be treated separately. As fairly general criteria for Exposition at a senior level you may find our categories helpful (pp. 51-52), and, if the modes selected at the intermediate level are similar to ours in the study, you may find our narrative and expository criteria helpful as well (pp. 22-24).

In the 1978 version of the project we employed Exposition: point of view in place of Exposition: process/procedure. These are three of the four criteria we shall be using for this mode (sentence structure remains as before):

Exposition: Point of View (Grade 8: 1978)

1. PLANNING

High
- *Evidence of a plan that is clear from the beginning.*
- *Ideas are developed in an appropriate sequence.*
- *A reasonable and clearly stated conclusion acting as a summary of the position.*
- *Logical points made with valid supports.*

Medium
- *Basic outline clear but sequencing illogical because key ideas are ineffective or missing.*
- *Too much space given to less important ideas.*
- *A conclusion that, if not satisfactory, makes a clear ending.*

Low
- *Unclear outline. Phases of the topic not adequately separated; contradictory or unrelated ideas may be found in the same paragraph.*

- *No clear definition of the writer's viewpoint.*
- *No clear conclusion or summation.*

2. *ARGUMENT*

High
- *Viewpoint clearly stated.*
- *Each main viewpoint substantiated with illustration, example, logic, analogy, contrast, or comparison.*
- *Some provisions made for alternative viewpoint with reasons supplied for holding that viewpoint.*

Medium
- *Viewpoint clearly stated.*
- *Some but not all main points substantiated as above. (High)*
- *No attention given to counterarguments; strongly one-sided.*

Low
- *The argument, such as it is, rests almost entirely on opinion.*
- *Few facts given in support.*
- *Poor relationship of facts or illustrations to what has been argued.*

3. *WORD CHOICE*

High
- *Appropriate level of formality consistent with subject, purpose, audience.*
- *Language choice precise and clear.*
- *Occasionally unusual or refreshing choice of words and images calculated to maintain interest on reader's part.*

Medium
- *As for High, except that the writer occasionally lapses into an unsuitable language register or dull trite phrasing.*

Low
- *Word choice is too vague or trite.*
- *Register of language unsuited to audience and purpose.*

The information concerning performance across the score range may be summarized in tabular form (as we have done in chapters 3 and 4, in graph form, or in both, depending on the audience and purposes.

Its main value lies in the information it provides that can be related to program: where the emphasis should be, where alternative teaching strategies might be considered, where present expectations about student writing at the grade level may be too high or too low. Its secondary, more general value, is to provide, at last, an empirical description of student writing as a counter to a wide miscellany of generalizations based on little more data than fond remembrance of the mythical good old days or the latest horrible example of writing quoted extensively in the newspaper. It is very important for teachers and the public to have sound, reasonably complete information for a change.

Though analytic scoring for positive characteristics of the writing appears "looser" than error counts, I believe it is the more valuable

analysis in characterizing writing at a particular level, and in providing information important for program decisions. Further, if a system is going after error counts (a popular topic), this negative approach to appraisal needs in fairness to all to be balanced by an approach emphasizing strengths.

Whatever else may be said of faults in the writing of grades 12-13 students, the evidence on this analysis is strong that students at this stage of their education do organize their writing reasonably well. The most serious weakness is the lack of colour or style both in language choice and sentence structure; this is the real challenge to writing programs in the Senior Division.

In comparison, the grade 8-level narrative, at least, shows relatively poor organizational control across three-quarters of the score range. This is a problem it appears reasonable to attack, perhaps in encouraging shorter, more pointed, narratives than may be the case in many classrooms presently.

Error Counts

As for the positive characterization, I recommend two scorers with scores averaged, especially where the decision has been made to note with some precision a number of narrow error categories. I am thinking here of various classifications of sentence sturcture errors and closely related serious punctuation problems. Pretraining is valuable, but no amount of it will bring scorers to total and consistent agreement on more subtle error classifications.

Some categories of errors may be worth counting in order to prove to sceptical audiences that in fact, in contrast to public consensus, this is not where the problems are. And that is important information both for teachers and the public at this time. For this reason, we went into more detail than I would necessarily recommend. I would not, for example, recommend teachers count capitalization errors (much too few) or diction errors (few errors and rather difficult to obtain consistency anyway). Despite their infrequency, however, it would be wise to count grammar errors related to the verb and to the pronoun. The frequency of "other grammar" was low, and likely unhelpful to program planning. Likely, counting "other grammar" is rather pointless: with the low error frequency no clearly defined problem area is likely to emerge, at least with

students for whom English is a first language.

At the other extreme, though, it may be important to point our repeatedly that "convention" errors far outnumber those in any class of grammar or sentence structure errors. I am not certain for all the work involved, that counts of spelling and minor punctuation errors will produce information of much use. A particular composition may be loaded with spelling errors; another with the same score may have none. Unless one were classifying carefully the errors made, I don't think the data would provide much information useful for general program decisions. A lot of students are poor spellers; a lot more are not. And, on our zealous counts at both grades 8 and 12 levels, I am prepared to concede that most students have many problems with minor punctuation.

Here, as in any other scoring or analysis procedure, the team within a system or school setting up the procedures for evaluating writing must decide which counts are, for the work entailed, likely to provide information of value in program decisionmaking.

The counts I recommend most highly concern various problems in sentence structure and (see below) the types of sentences used. Here I do suggest breaking down the classification "Sentence Structure" or "Major Sentence Structure" into specific error types: "no sentence" (happily infrequent), "run-on," "fused," and so on.

Combining the analytic appraisal of sentence structure and error count information, I have observed in the writing in our studies that most students make relatively few serious errors, but that most do not use sentence style for effect. It is here that I would place more program emphasis with attention to contemporary models of writing and to imaginative sentence-building exercises.

Concerning the high frequency of punctuation errors, a finding I am sure will be reinforced by other studies, I do not recommend a rule-based approach. Many punctuation errors I found in personal essay-by-essay examination of the writing are the result of not "hearing" the sentence, of a lack of sensitivity to its intended effect. The attack on the problem seems best directed through a constructive examination of good sentence style as well as teaching the student to listen for sentence rhythm and natural pauses.

Measures of Syntactic Maturity

Few of the measures we used appear to have much utility for examination of performance at a single grade level. If your evaluation is restricted to a single grade level, I would not recommend the T-unit measure or measurement of average sentence length. The counts I would recommend concern the kinds and variety of structures of sentences. What is the frequency of an introductory clause or phrase? What is the frequency of compound sentences? Or, perhaps (a count we didn't make), what connectives and sentence adverbs are most frequent? This provides some useful data on the present state of the students' mastery of sentence style, data useful in planning units for program and for later checks on the effectiveness of such units in improving performance.

If the evaluation plan includes the study of writing at two non-adjacent grade levels or at several grade levels, the T-unit measure has significance. It is evidence of maturation of writing style and may be of some assistance, over time, in determining whether a systematically applied program change has had an effect. I should add that the T-unit count is straightforward; you will need only one scorer per essay for this and the other measures mentioned. The definition of the T-unit, with illustrative sentences, is given on p.4.

SUMMARY

The first and critical step for the planned evaluation of writing within a school or a system is the establishment of reasonable consensus of expectations concerning the writing program at the grade levels concerned: these expectations include careful identification of the modes of writing reasonable for the grade and of a number of criteria reasonably applicable to these modes.

I strongly recommend that teachers draw on available published resources to a considerable extent in defining criteria so the task of specifying criteria will not be too burdensome.

The second step is a reasonable standardization of the manner in which the student's writing is kept on file together with appropriate records. For this, I recommend the writing folder and frequent use of analytic scoring by teacher and student.

The third step, assuming one wishes to include a reasonable sample of writing over the term, is to achieve some agreement about the frequency,

modes, and scoring treatment of formal assignments over the school year, with the understanding that scores on these will form a substantial portion of the final mark.

The fourth step is decision concerning the final pair of assignments for holistic scoring and the conditions under which they will be written and scored. Mode and topic must be controlled.

This decision should be made early in the school year and all information except the specific topics shared with the students.

At this stage also a number of organizational matters - type of paper, time and locations for writing, student coding, distribution of essays for scoring, reasonable concessions to scorers, and compilation of data - should be dealt with. The plan should permit students to receive their scores before the end of the school year.

The fifth step is the analysis of a random sample of the writing by a small trained team. The dimensions of that analysis should be agreed to early in the project with consideration of the value of certain kinds of information to different audiences or "clients" and very particular consideration of the value of the information in making program decisions and assessing the effectiveness of program change.

Bibliography

1. Important Resources for Teachers and Consultants

Britton, J., T. Burgess, N. Martin, A. Mcleod, H. Rosen, *The Development of Writing Abilities*, Macmillan Education, London, 1975 (available from the National Council of Teachers of English).

This study presents a new and more realistic set of categories to classify what writers actually do in writing. The analysis looks at writing from two angles: The function served by a piece of writing and the audience to which it is addressed.

Cooper, C. R., "Measuring Growth in Writing," *English Journal*, vol. 64, no. 3 (March 1975).

This summary discusses briefly some of the problems in evaluating writing, some of the measures that should be considered in evaluating student writing, and a scheme for whole-school measurement of growth in writing.

Cooper, C. R., L. Odell, *Evaluating Writing*, National Council of Teachers of English, Urbana, 1977.

A comprehensive summary of current information on describing writing and measuring growth in writing.

Diederich, P. B., *Measuring Growth in English*, National Council of Teachers of English, Urbana, 1974.

The writer discusses various problems of evaluation in English, and suggests a criteria rating system for a school to undertake.

Evans, P., M. Marsh, P. Brown, *Criteria for the Evaluation of Student Writing: Grades 7 & 8. A Handbook*, Ontario Institute for Studies in Education, Toronto, 1977.

A list of various criteria for several modes of writing that might be used in evaluating student writing in grades 7 and 8.

Hunt, K. W., *Grammatical Structures Written at Three Grade Levels*, National Council of Teachers of English, Champaign, 1965.

This study was a search for developmental trends in the frequency of various grammatical structures written by students of average IQ in the fourth, eighth, and twelfth grades. The T-unit is defined and used in the analysis.

Mellon, J. C., *National Assessment and the Teaching of English*, National Council of Teachers of English, Urbana, 1975.

A good summary and a good commentary on the National Assessment of Educational Progress's assessments of knowledge, abilities, and attitudes in English.

Moffet, J., *Teaching the Universe of Discourse*, Houghton Mifflin, New York, 1968.

An excellent analysis of the development of language and writing in children. Suggestions for building "a student-centred language arts curriculum."

National Council of Teachers of English, *Common Sense and Testing in English*, National Council of Teachers of English, Urbana, 1975.

Focusses on legitimate as contrasted to illegitimate use of test scores. Good advice to teachers on test selection.

O'Hare, F., *Sentence Combining*, National Council of Teachers of English, Urbana, 1973.

This study suggests sentence combining practice as a way to improve not only the syntactic maturity of students' writing but also the overall quality of such writing.

2. Other Significant Studies in Writing Evaluation

Braddock, R., R. Lloyd-Jones, L. Schoer, *Research in Written Composition*, National Council of Teachers of English, Champaign, 1963.

Britton, J. N., N. C. Martin, H. Rosen, *Multiple Marking of English Compositions*, Schools Council Study no. 11, Her Majesty's Stationery Office, London, 1966.

Cooper, C., A. Purves, *A Guide to Evaluation*, Xerox, 1974.

Department of Education and Science, United Kingdom: Report of the Committee of Inquiry...under the Chairmanship of Sir Alan Bullock, *A Language for Life*, Her Majesty's Stationery Office, London, 1975.

Evans, P. J. A., "The Intermediate Writing Evaluation Project," *Orbit 42*, vol. 9, no. 2 (April 1978).

Follman, J. C., J. A. Anderson, "An investigation of the reliability of five procedures for grading English themes," *Research in the Teaching of English*, vol. 1, no. 2, 1967.

Loban, W. D., *The Language of Elementary School Children*, National Council of Teachers of English, Champaign, 1963.

Loban, W., *Language Development: Kindergarten through Grade Twelve*, National Council of Teachers of English, Urbana, 1976.

Mellon, J. C., *Transformational Sentence Combining*, National Council of Teachers of English, Champaign, 1969.

O'Donnell, R. C., W. Griffin, R. Norris, *Syntax of Kindergarten and Elementary School Children: A Transformational Analysis*, National Council of Teachers of English, Urbana, 1967.

Report 8: Writing Mechanics, National Assessment of Educational Progress, Denver, 1972.

Report No. 05-W-01, Writing Mechanics, 1969-1974, National Assessment of Educational Progress, Denver, 1975.

Report No. 05-W-02, Expressive Writing, National Assessment of Educational Progress, Denver, 1976.

Report No. 05-W-03, Explanatory and Persuasive Letter Writing, National Assessment of Educational Progress, Denver, 1977.

Russell, H. H., C. Wolfe, P. Evans, R. Wolfe, R. Traub, A. King, *Programs and Student Achievement at the Secondary-Postsecondary Interface: Interproject Analysis*, Ontario Ministry of Education and Ministry of Colleges and Universities, Toronto, 1976.

Stewart, M. F., "Syntactic Maturity from High School to University: A First Look," *Research in the Teaching of English*, vol. 12, no. 1 (February 1978).

Traub, R., R. Wolfe, C. Wolfe, P. Evans, H. H. Russell, *Secondary-Postsecondary Interface Project II: Nature of Students*, vols. 1 and 2, Ontario Ministry of Education and Ministry of Colleges and Universities, Toronto, 1976.

Address for NCTE:

National Council of Teachers of English
1111 Kenyon Road
Urbana, Illinois 61801
U.S.A.

Appendix

EXEMPLAR ESSAYS: INTERFACE STUDY

From pp. 71-89 of Traub et al.: *Nature of Students*, Vol. 2

Following my informal analytic scoring of a sample of the essays in the Interface Project, I selected six essays, not to typify student writing, but to illustrate some of the trends in writing quality across the score range. Though in the present study we have carried out a more formal appraisal, the reader may find these "exemplar" essays and my earlier comments of some help in giving some colour to the tabular information and relatively brief comments in chapter 4, Part 3 of the present report. Comments on errors are excluded from the following appraisals as the sequence was intended to illustrate features of the "positive characterization" of the writing.

FIRST EXAMPLE (Score 3.7)

Topic: Secondary schools should not demand that students take particular courses or prerequisites; students should be free to choose courses of interest to them.

The debate of whether students should have a choise by cources they can take in high school, can have two sides. On the pro side, there would be most students who are about to entre or have just entered the secondary level of education. The reason for this is quit simple, people want as much freedom as they can get. I know, several years ago as a student in grade eight, the idea of freedom of choise was quite appealing. But today, I can see how easily it could of been for me to make a serious error in course choises. This occured to me only after it was to late to make any alterations. Forcunatly, though, I did make the right choices, and therefore, I can continue my education as I wish. But other people I know have not been so fortunate. They made the wrong choices when they first entred high school, and today they cannot follow

the course of education they wish. The main reason for this occurance is that students are made to make this desision when they are quite young and do not fully realize the importance of the desision.

On the con side of the arguement are the older people or people who have finished thier education. These people on the whole would like to see the cources pre-set and compulsory to all students. In this way, they make the assumption that this will result in better educated students to enter university. This is a certain degree can be a valide point. But there is one falicy to the arguement, students on the whole will do better in subjects they enjoy taking than in subjects that do not perticularly appeal to them. I know personaly that subjects I enjoy taking average as much as twenty percent higher than subjects I don't perticularly enjoy. This I emagine can be said for all students. The result therefore is that if you make students take many subjects that they do not enjoy, then they will not work as hard at them and thier marks will generaly be lower than if they had choosen the course themselves

At this point, it may appear that I am sitting on a fence and not making any altenative methods. But I feel that a compromise can be made between the pro and the cons. I feel that some cources should be compusory. These subjects are English, mathematics, History and science. English is an important pre-requisite since almost all university require english before they will accept graduates. In addition, english is a subject that makes the backbone of a persons intelegence. If a person has many good ideas and thoughts, he can be considered inteligent, but if he cant express his ideas in an intelecuale fasion, then he may not appear to be an intelecual.

Mathematics is an important subject becaus for graduates to get into university under various cources in sciencses and engenering and medicine, he may need mathematic cources from secondary schools.

History is a necessary cource for students to take. Of all the cources you take at high school, history has one of the most practical uses in every day life. If a person ever gets involved in a discusion, the knolage of historical events may be very useful in defending a point.

Science is also an important course that should be taken, at least in the level one grade, because it gives the student a better understanding of the world around him.

I feel that all the cources I have stated above should be taute at the advanced phase level until the end of the level two year. In this way students are much more aware of their own goals they want to take, whether it being university or colages of applied arts and technology. After the level two year, I think that a greater freedom of choice of cources should be given with english and mathamatics being compulsory.

Therfore, if a compromise between the two sides is drawn with the outline above, the level of education will be bettered without the expen [illegible] *of the freedom of the student being taken away.*

COMMENTS ON FIRST EXAMPLE

This essay, though replete with mechanical errors, illustrates the point that essays relatively low on the scale did show some effectiveness in organization and paragraphing. It follows a "Here's one side; now here's the other" pattern, followed by development of the writer's own position in some detail and with reference to several subjects the student considers important. And even if the conclusion is limp, it does provide a summary statement as the student steps back from particulars to a final generalization.

No doubt paragraphing could be more effective; a better design might have been to deal with specific subject areas in fewer paragraphs, but there is at least a sense of "one idea—one paragraph."

The tendency to weak generalizations is illustrated by this essay: "On the con side of the argument are the older people...(who) would like to see the courses pre-set and compulsory to all students," as if there were clearly two camps of the human race—entering Grade 9ers and "the older people"—the one group holding one position universally, and the other, the other. What the student refers to as the fallacy of the latter position (that "I," and likely most other students, do well in courses I like) ignores the possibility that students tend to "like" the courses they do well in (i.e. have the capacity for). On the positive side, the student does attempt to be specific in reference to his own experiences and in noting specific subjects. A reasonable sequence of reasoning can be followed and there is balance in the argument.

In style, both sentence structure and diction, the essay is wholly undistinguished, characterized by a lack of terseness and little consciousness of word values. For example, the writer adds an explanatory sentence,

"The main reason for this occurence is that students are made to make this decision when they are quite young..." rather than attaching the explanation to the previous sentence with a linking phrase such as "chiefly because."

The essay has some flavour: one senses a commitment to the position taken and evidence of deliberateness in its development. The personal note is frequent and desirable; without it the essay would be extremely dull.

SECOND EXAMPLE (Score 5.5)

Topic: Should scenes of violence in books, in films, and on television be censored?

It is hard to say whether cencorship should exist at all or to what degree. If everyone was of sound-mind and basically a good character we wouldn't have to worry about it. Evidently however even all us Canadians don't qualify for the sound-mind bit. Unfortunatly also not all the ones deserving a rubber room have one either so were left with a lot of good people a few harmless crazy ones and a few who aren't so harmless.

I've read a lot of books and enjoyed most of them. I'm just trying to think of them without their crime festered plots. They'd probably be really dull and boring for most people. Maybe it would prohibit one idea of the perfect crime but Canadian illeteracy % would likely skyrocket.

Last year, several schools throughout Canada and the United States were plagued with shooting incidences. A mentally disturbed student would enter a school and open-fire in the halls and classrooms. Many people were killed and everyone seriously frightened.

Undoubtedly, the tremendous amount of coverage on the topic and the general outroar from society against this particular crime somewhat encouraged these students. It really makes you wonder whether were not all to blame. At the beginning of each new television season we witness more and more crime. Our need for gory deaths and drastic amounts of blood never seems to be fulfilled. Sometimes changing the channel doesn't help much either. Our taste for good "shows" is becomming warped and contorted -we like crime! -

However, I'll admit we're humane. Sure enough after the above mentionned killings new laws prohibiting certain guns and rifles were introduced. How well they are enforced and by whom is up to us. We really can't expect too much if we think well you know—my son can have a rifle- he's got it all upstairs!

Canadian crime statistics have increased over the past few years along with the population. Provincially Ontario came out number one again and not by a small persentage.

If you were to ask a father what crime he considered the worse, he'd likely tell you, rape. thinking back tell September, I can recall hearing about or seeing at least twenty shows involving rapes. Is it so odd then that the number of rapes is also ascending? Whether it be good old-fashioned Dr. Welby, Medical Centre, Policewoman or movies, we cannot escape the present trend. We'd all likely vote rape in our top five most dreaded crimes for women and we'd disown a raper son.

It use to be, a women had to prove her innocence when she pressed rape charges. Her history was gone over and if possible she was made to look like a real tramp. The raper was presented to be the all-american boy. Recently however, a new bill was passed whereby a woman's history would not be revealed when dealing with this sort of crime. Instead the pendulum swings the other way and the raper must try to prove his innocence. More of this should be shown on televison. It might serve to discourage would-be rapers if they thought their chances of getting caught were a lot higher.

I'm just not content to say well were not as bad as the States!

Presently Judy LaMarsh is heading a committee in charge of censoring crime on T.V. I think it is a worthwhile cause if done properly. Right now, I find Canadian T.V. censorship too lax. For example "Deliverance" was on T.V. not too long ago, the next day I saw one little boy grab the ear of another kid and told him to squeal like a pig!

Everyone knows that even little kids don't go to bed right after dinner so shortly after T.V. family hour doesn't really make a difference in the age of the audience.

I find that many movies shown at theatres which are resticted have less to offer in the way of crime as do many T.V. shows.

I agree censorship should exist but to certain degrees.

COMMENTS ON SECOND EXAMPLE

There is an organizational thread running through this essay, though the writer's view of the topic does not become especially clear until late in the essay, because he (she) jumps out early into a number of sociological dimensions, presumably effects of excessive TV violence, before coming back to state the argument or case. There is, then, a plan, though one that could be made more effective.

Here, however, we do see improvement in the effectiveness of argument, though the essay remains somewhat disjointed and some of the assumed causal connections are rather far-fetched (just as they often are in magazine or newspaper articles on the censorship of violence). He is reasonably concrete in his illustrations: note especially, paragraph ten. And among his reflections emerges the possibility that it is not our television that is to blame, but ourselves - the society and what it values. There are insights that prevent him from making blanket generalizations or suggesting simplistic solutions.

There appears to be a genuine attempt to vary sentences for effect, perhaps an over-reliance on exclamation marks. Diction is uneven in quality; sometimes words are chosen with precision, at other times the choice is overly colloquial and there are actual errors. (Or is "outroar" an interesting coinage?)

Though there are many matters to fault in the essay, the writer has compelled the reader's interest with some touches of humour (even if rather trite) and with his directness. Through the writing we catch something of the writer as a person, a feature rather infrequent across the sample of essays examined. The essay does indicate that "flavour" is not a quality of writing exclusive to the best essays.

THIRD EXAMPLE (Score 6.0)

Topic: What are the values of the Olympic Games, not only for the competitors, but especially for the countries they represent and for the world?

In this essay I will be discussing the values in the Olympic Games, not only for the competitors, but especially for the countries they

represent and for the world.

The year 1976 is an important one for Canada. This year, in Montreal, Canada has the honour of holding the summer Olympics which will commence in July. The Olympic Games have had excellent results in their money raising campaigns through the Olympic lottery. This is one source of money for the construction of the buildings many will see this summer. The second source of money, unfortunately comes from the Montreal taxpayer's pocket.

The Olympic's value in a country is the union it brings. Every year either in summer or winter the games that are held in a chosen country unite all countries of the world. Even if there are four athletes present to participate in games there is no limit to the amount of persons that should represent their country.

The Olympics this year in Canada will bring the awareness of how important sports are for Canadians. Millions of Canadians are not fit, hopefully the Games will motivate them to take up one of the numerous sports offered.

The Olympic Games also bring in tourists from all the corners of the world. This not only provides wealth for the country, particularly the city hosting the games but also provides an image of beauty for a country. Once the visitors come to a country that shows warmth and friendship as seen in Canada there will be a want for return or a recommendation to their friends.

For those Canadians who do not attend the Olympic Games, they will be able to see the buildings in Montreal whenever they wish. The buildings in Montreal: the veladrome, the centre arena, and the Olympic Village will remain there, in memory of the honour, beauty and union that the Olympics once offered in Canada.

The Olympics for Canada brings a union within the country. In recent years there has been a major problem of bilingualism. This is heard in the province of Quebec. They feel Quebec's major language should be French which has faded away in the last couple of years. The Games will provide a unity for the east and west side of Canada.

The value the competitors take in the Olympic Games is the most exciting experience in their life. The feeling of representing their country in seen all over their face. Even if the athletes do not gain a

medal they are not sore losers but feel they gave it a good try and besides everything else it was worthwhile. The athletes of each country get to communicate and learn about other cultures of their competitor. There is friendship between two countries shown by the athletes.

The athletes show the importance of a particular sport in their country. This is shown by their training, skills, and a medal usually. For example: In West Germany, sports play a major role. After work every individual has a particular activity they take up. This shows the importance of fitness in that country. Athletes work very hard for long hours day after day. They eat certain foods, and not too much entertainment is allowed on weekdays. Almost everything is under strict control. These results lead to the other country's awareness of a higher aim for the next year. Either winners or losers all competitors rate the same to each other. They know due to experience the hardship one must go through to enter the Olympic Games.

Therefore we may conclude that it is through the competitors and the country as a whole that bring out the values of the Olympic Games.

COMMENTS ON THIRD EXAMPLE

This essay, though reasonably adequate in organization, illustrates the general blandness of the writing in the middle of the range. Viewed from the standpoint of errors, mechanical and other, this essay appears to be more effective than it really is.

The naiveté of the argument and lack of substantive support are obvious at once; views of the Olympics as a cure for bilingual problems, an expression of world friendship authored by the host country, and an event with real impact on sports and fitness all seem rather futile (though the reader must remember that the essay was written before the Games). The writer has "covered" the topic, but thinly, and with little sense of direction or purpose. The concluding sentence is limp — just a sign-off.

There is little awareness of sentence style (note the choppiness of the many sentences in the ninth paragraph), and diction and imagery remain wholly abstract.

The blandness of effect is explained and reinforced by the absence of flavour; there is no sense of personal commitment or interest. The one

touch of humour (last sentence of paragraph two) stands out only because nothing else in the essay does.

While many commentators on student writing at the interface have been greatly concerned about error frequency, the frequency of such apathetic writing perhaps should be a matter of even greater concern. Doubtless the test circumstances were a significant contributing factor in that they did little to stimulate writer commitment; nevertheless, the dryness of so many essays suggests that more might be done in schools to encourage commitment and vigour in writing. The student with a stake in his ideas and with a richer sense of audience is the one who is likely to write purposefully and effectively.

FOURTH EXAMPLE (Score 6.8)

Topic: Should scenes of violence in books, in films, and on television be censored?

Have you read Stranger in a Strange Land by Robert A. Heinlein. If not, get up off your ass and go buy a copy. On second thought stay there and I'll tell you the pertinent point. (How's that for alliteration?) Heinlein says that man is the only creature that laughs and that mankind laughs "because it hurts". This is a very emotional world with an abundance of pain, both physical and mental. Is pain an emotion? A feeling? No, it is pain; intangible but ever present. Pain is a guillotine over your head that is ready to drop at any instant. Pain hovers above you at all times and each one of us becomes changed in some way due to our efforts to avoid its presence.

Does violence warp a child's mind? Does he become "strange" because of his efforts to avoid or perhaps even seek the pain he has been exposed to? I was weaned on violence and look how well I turned out. Or did I? Now turn around and look at the occupants of the world's prisons and asylums. And graves. Most of those people weren't born that way: They were changed, warped by society. Somewhere along the line they were twisted by their environment.

It is time to try a little preventative medicine. We can never abolish violence but we can try control the market we sell it to. Go ahead and censor the violence on television during the hours that children watch. Censor that idiot box while you're at it. But remember that

if you initiate your censorship program you will put a lot of industries out of business. Certainly the cop shown would survive: there will always be an adult market for such trash. But what about the cartoons and comic books. Even Walt Disney films have violent scenes. Those comic books are really incredible: "Triple A rating - No sex, no drugs, no foul language", just picture of three mice going through a meat grinder and a cat swallowing a bundle of dynamite. When was the last time you watched Bugs Bunny? Did you see him shoot Elmer in the face with a shotgun? Funny eh? Have ever laughed at a friend who just fell on some slippery ice? Don't deny it, you thought it was hillarious. You laughed because it hurts. The Three Stooges built a comedy empire on pain and violence. Think of the last good joke you heard. Chances are that if the joke wasn't dirty it's meaning was actually rather sad. If you eliminate violence you will eliminate a great deal of laugher too.

I guess you realize that I haven't taken a firm stand on this issue. I believe that we would solve a lot of emotional problems, clear a lot of mental wards, if we were able to stop teaching violence to impressionable youngsters. But I can't be sure that those children would be able to cope with the violence they are sure to meet in the "jungle". I'm not sure that I could stand to see people living on half-rations of laughter either.

I'm laughing now because it sure as hell hurts.

COMMENTS ON FOURTH EXAMPLE

This essay is included to illustrate the observation made in the introductory comments that some of the most interesting and flavourful writing was exhibited where the student refused to take the test too seriously. The writer-reader relationship changes dramatically from a dreary "writer to examiner" style, illustrated in the previous essay, to more of a "writer to his peers" style.

Probably because this student has broken out of the mold, he provoked a wide variety of responses from the scorers who rated the essay all the way from 3 to 10 on the scale. Apart from the mechanical faults, evidently some scorers reacted strongly and negatively to abrupt shifts into a language register generally colloquial and of questionable appropriateness. Other scorers likely over-reacted in the opposite direction, simply out of relief at the gasp of fresh air this writer provided. No

one can say the essay is weak on flavour, even if the flavour is too strong for many.

Leaving aside "register," there is an unevenness in the writing and argument. The quasi-philosophical reflection on pain seems partly irrelevant. And flippancy is frequently indistinguishable from mere carelessness. There is a logic to the essay but the personal intrusions and illustrations, even if germane, are introduced in a way that partly obscures the direction of thought.

On the other hand, many of the jabs are good ones. Are not the observations on contents of comic books sharp and fair? And the writer is not the first observer of the human scene to note the interplay of pain and laughter. On a second reading, it is clear that this student operates well in mingling the concrete and specific with the universal.

Diction, barring the "bottom of the vernacular" descents, is generally precise if not outstanding, and certainly the writer uses sentence style to create effects.

A quick and vigorous mind is at work here, revealing itself in somewhat heterodox ways. And even if these are perceived as somewhat extreme, the essay suggests that if a more effective stimulus were provided, one that placed writer and intended reader in a more intimate relationship, rather different and much more interesting qualities of writing might emerge.

FIFTH EXAMPLE (Score 8.0)

Topic: Does art imitate life, or does life imitate art?

A piece of art that imitates life is not a true work of art. A person that lives his life imitating art is not truly alive. Art and life are not carbon copies of each other, but rather reactions to experiances.

When a author sets out to write a book he doesn't regard events. That is the reporters job. Instead, the author writes about reactions to the events, from inside him rather than from outside. The artist who paints a picture does not try to imitate the scene. A camera will serve that purpose. The artist will attempt to record his reaction to the scene presented. The ballet dancer can perform the steps perfectly but

unless his or her own interpretation of the steps is added, the performance is dead, simply an imitation of life. We have examples of the limitations. Every corner book store has that one shelf of books Mother forbade you to look at. These are imitators and can hardly be called "art." The difference between true art and an imitation of life is the difference between DaVinci and Playboy; between Robert Wagner and Alice Cooper; or between D. H. Lawrence and the man who sits in his attic apartment spewing out 5 book a week.

Life, like art, is a series of sponteneas reactions. By attempting to imitate anything, especailly art, we inhibit these reactions and become only half a person, the biological part. What makes us human is not only our physical selfs but also our spiritual beings. Just as our body would reject an imitation, any attempt to supply imitations to the spirit would also be rejected. Living a life that attempts to imitate art would be like gazing at a photograph. Oh, the pictures are very nice but everyone is a bit to stiff and unnatural, the smiles, a trifle to wide.

As a carbon copy is fainter than the original, an attempt to imitate life as art would never be true. Life and art are similar as they are both reactions but reactions can never be recorded and replayed. Here today, gone tommorow is what makes life and art what they are.

COMMENTS ON FIFTH EXAMPLE

This topic, not a very popular one, produced some very interesting essays. Those most interesting were the ones where the students attempted, as in the present example, to carry mental explorations somewhat further than maturity and knowledge warranted: truly an "essai" or a "try." Generally scorers treated even relatively unsuccessful "tries" with generosity.

This essay illustrates well the quality of essays placed near the top of the score scale. Though the essay is short it is characterized by tautness of language and organization. Statements of position, even if somewhat erroneous in fact, are put forward bluntly and unambiguously. This directness and absence of padding gives a freshness of flavour so frequently lacking in other essays and the choice of language is generally appropriate to the subject matter.

Argument is well developed by use of contrast (life: art; artist: reporter; body: spirit) and use of illustration.

That the student is reaching beyond himself is apparent from his conclusion; he is out of his depth and does not know quite how to reach shore. The reader must, however, admire the swim.

In this essay, we see a more mature effort to relate sentence structure to purpose. If the writer had learned the semi-colon, he might have used ellipsis and antithesis to better effect. However there is a good rhythmic sense and usually variety in structure relates to effect intended.

SIXTH EXAMPLE (Score 9.2)

Topic: A competitive environment develops strength of character.

Today's society is to a great extent a competitive society. Early in life we are taught the importance of "getting ahead" and "striving to be better than the neighbour." Throughout our school system, "success" is measured directly according to one's ability to achieve "more" than one's classmate. When looking for a job, one is forced to compete in order to make an impressive impression. Indeed, in a capitalistic society such as our own, the "competitive environment" is a predominent factor. Some people see no harm in this: philosopher Thomas Hobbes insisted it was "in man's nature" to compete.

Is this so? Or does society force man to be competitive? And if so, does this competitive environment develop strength of character?

Because of the continuous indoctrination of competitiveness throughout one's early life, it is difficult to lose the competitive nature. As a result, men are forever "looking out for themselves"; trying to better themselves; trying to "get ahead." In a capitalist system, the ultimate result is the exploitation of man by man, of class by class. Individuals strive for personal, material gain. Instead of developing strength of character and humane values, men become obsessed with making money. Because our society is competitive, and because it places a high value on material wealth, it naturally follows that men strive to further personal wealth. The "profit motive" overrides the importance of humane values.

As a result, those who are unable to "compete" live often in worlds of mediocrity or even poverty, while those who are successful competitors live in affluence and are looked up to by society. These "successful competitors" feel little or no obligation to enhance the living conditions

of their fellow men. The competitive environment in which they have been brought up has conditioned them into believing that they owe nothing to their "competitors". There is no consideration for the needs of those who are less fortunate.

Instead of building a strong character - one which has consideration for others - the competitive environment conditions man to be narrow-minded and concerned only with self gain. Moreover, the competitive environment creates hardships for those who are unable to compete. Because of our society's overemphasis of the competitive environment, many children go hungry and many elderly citizens are unable to live with dignity. Hopefully, our society will recognize this fact soon, and transform it from a competitive society to a society concerned with brotherhood.

COMMENTS ON SIXTH EXAMPLE

This essay represents the top of the scale as available to us in the sampled essays. It requires very little comment.

Like the previous example, it is tautly constructed; there are no waste words or irrelevent ideas. Plan and direction are always clear.

In argument the student articulates his position clearly; however, as in most of the essays, supportive evidence or illustration is not available. In this essay, the logic of the case is there and the care for development has the reader nodding rapidly in agreement rather than looking for evidence. He would have to look hard, for the essay is built on generalizations about "people," "man" or "men," "some people," and "our society." The student provides a definition of a "a strong character" which is rather noble, but it is not a definition that would win universal agreement and it is not supported.

Flavour is good simply because the student feels strongly and expresses himself effectively. The reader, whether he agrees with the position or not, knows whom he is listening to.

Sentences are varied effectively in relation to purpose. The use of a series of questions in a short paragraph is a very effective means of shifting from description to argument. The periodic sentence with the "lead-in" by means of subordination, if somewhat overworked, is generally effective, especially in development of the argument sequence.

The level of diction, if not exceptional, is generally suited to the topic, but marred by the student's tendency to italicize (by quotation marks) too often.

As in many of the essays, the conclusion is limp. In this instance, as in the previous example, the exaltedness of theme leaves the student floundering for an adequate conclusion.